

Cambridge Studies in Linguistics 128

Grammatical Categories

Variation in Romance Languages

M. Rita Manzini and
Leonardo M. Savoia

CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521765190

This page intentionally left blank

GRAMMATICAL CATEGORIES

Grammatical categories (e.g. complementizer, negation, auxiliary, case) are some of the most important building blocks of syntax and morphology. Categorization therefore poses fundamental questions about grammatical structures and about the lexicon from which they are built. Adopting a 'lexicalist' stance, the authors argue that lexical items are not epiphenomena, but really represent the mapping of sound to meaning (and vice versa) that classical conceptions imply. Their rule-governed combination creates words, phrases and sentences – structured by the 'categories' that are the object of the present inquiry. They argue that the distinction between functional and non-functional categories, between content words and inflections, is not as deeply rooted in grammar as is often thought. In their argumentation they lay the emphasis on empirical evidence, drawn mainly from dialectal variation in the Romance languages, as well as from Albanian.

M. RITA MANZINI AND LEONARDO M. SAVOIA are both Full Professors of General Linguistics at the University of Florence.

In this series

94. JOAN BYBEE: *Phonology and language use*
95. LAURIE BAUER: *Morphological productivity*
96. THOMAS ERNST: *The syntax of adjuncts*
97. ELIZABETH CLOSS TRAUGOTT AND RICHARD B. DASHER: *Regularity in semantic change*
98. MAYA HICKMANN: *Children's discourse: person, space and time across languages*
99. DIANE BLAKEMORE: *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*
100. IAN ROBERTS and ANNA ROUSSOU: *Syntactic change: a minimalist approach to grammaticalization*
101. DONKA MINKOVA: *Alliteration and sound change in early English*
102. MARK C. BAKER: *Lexical categories: verbs, nouns and adjectives*
103. CARLOTA S. SMITH: *Modes of discourse: the local structure of texts*
104. ROCHELLE LIEBER: *Morphology and lexical semantics*
105. HOLGER DIESSEL: *The acquisition of complex sentences*
106. SHARON INKELAS AND CHERYL ZOLL: *Reduplication: doubling in morphology*
107. SUSAN EDWARDS: *Fluent aphasia*
108. BARBARA DANCYGIER AND EVE SWEETSER: *Mental spaces in grammar: conditional constructions*
109. HEW BAERMAN, DUNSTAN BROWN AND GREVILLE G. CORBETT: *The syntax-morphology interface: a study of syncretism*
110. MARCUS TOMALIN: *Linguistics and the formal sciences: the origins of generative grammar*
111. SAMUEL D. EPSTEIN AND T. DANIEL SEELY: *Derivations in minimalism*
112. PAUL DE LACY: *Markedness: reduction and preservation in phonology*
113. YEHUDA N. FALK: *Subjects and their properties*
114. P. H. MATTHEWS: *Syntactic relations: a critical survey*
115. MARK C. BAKER: *The syntax of agreement and concord*
116. GILLIAN CATRIONA RAMCHAND: *Verb meaning and the lexicon: a first phase syntax*
117. PIETER MUYSKEN: *Functional categories*
118. JUAN URIAGEREKA: *Syntactic anchors: on semantic structuring*
119. D. ROBERT LADD: *Intonational phonology second edition*
120. LEONARD H. BABBY: *The syntax of argument structure*
121. B. ELAN DRESHER: *The contrastive hierarchy in phonology*
122. DAVID ADGER, DANIEL HARBOUR AND LAUREL J. WATKINS: *Mirrors and microparameters: phrase structure beyond free word order*
123. NIINA NING ZHANG: *Coordination in syntax*
124. NEIL SMITH: *Acquiring phonology*
125. NINA TOPINTZI: *Onsets: suprasegmental and prosodic behaviour*
126. CEDRIC BOECKX, NORBERT HORNSTEIN AND JAIRO NUÑES: *Control as movement*
127. MICHAEL ISRAEL: *The grammar of polarity: pragmatics, sensitivity, and the logic of scales*
128. M. RITA MANZINI AND LEONARDO M. SAVOIA: *Grammatical categories: variation in Romance languages*

Earlier issues not listed are also available

CAMBRIDGE STUDIES IN LINGUISTICS

General Editors: P. AUSTIN, J. BRESNAN, B. COMRIE,
S. CRAIN, W. DRESSLER, C. J. EWEN, R. LASS,
D. LIGHTFOOT, K. RICE, I. ROBERTS, S. ROMAINE,
N. V. SMITH

*Grammatical Categories: Variation
in Romance Languages*

GRAMMATICAL CATEGORIES

VARIATION IN ROMANCE
LANGUAGES

M. RITA MANZINI

University of Florence

LEONARDO M. SAVOIA

University of Florence



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City
Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521765190

© M. Rita Manzini and Leonardo M. Savoia 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Manzini, Maria Rita

Grammatical categories : variation in romance languages / M. Rita Manzini,
Leonardo Maria Savoia.

p. cm. – (Cambridge studies in linguistics ; 128)

ISBN 978-0-521-76519-0 (hardback)

1. Grammar, Comparative and general—Grammatical categories. 2. Language
and languages—Variation. I. Savoia, Leonardo Maria, 1948– II. Title.

P240.5.M36 2011

415–dc22

2010052183

ISBN 978-0-521-76519-0 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs or external or third-party internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate.

Contents

<i>List of tables</i>	<i>page</i> x
<i>Acknowledgements</i>	xi
Introduction: grammatical categories and the biolinguistic perspective	1
1 The structure and interpretation of (Romance) complementizers	13
1.1 Romance complementizers are nominal and head their own noun phrase	14
1.2 Structure of the complementizer phrase	19
1.2.1 Combining a left periphery in the complementizer phrase and in the embedded sentence; combining two complementizers	23
1.2.2 Some potential problems	30
1.3 The left periphery beyond complementizers	37
1.3.1 Is order dictated by interpretation – or interpretation by order?	38
1.3.2 Embedded contexts	43
1.4 Conclusions	47
2 Variation in Romance <i>k</i>-complementizer systems	49
2.1 Systems with two <i>k</i> -complementizers	49
2.1.1 Definite and indefinite complementizers – and alternative analyses	54
2.1.2 Generalized <i>wh</i> -complementizers	61
2.2 ‘If’	65
2.3 The interaction with (non-)finiteness	73
3 Sentential negation: adverbs	80
3.1 Sentential negation adverbs are nominal and argumental	83

3.1.1	Further evidence	88
3.1.2	Sentential negation adverbs as nominal arguments	92
3.2	Ordering sentential negation with respect to other adverbs	94
3.2.1	The order of negation with respect to aspectual adverbs	106
3.2.2	The order of negation with respect to quantificational and manner adverbs	111
3.2.3	General discussion	118
3.3	The interaction of adverbial and verbal positions: the participle	120
4	Sentential negation: clitics	128
4.1	Interactions of negation clitics and subject clitics	131
4.2	Interactions of negation clitics with object clitics	138
4.2.1	Non-negative <i>n</i>	145
4.3	Negative concord and negative doubling	152
5	The middle-passive voice: evidence from Albanian	159
5.1	Data	160
5.1.1	Middle-passive morphologies	160
5.1.2	The interpretation of the middle-passive morphologies	164
5.1.3	The Arbëresh varieties	169
5.2	The <i>u</i> clitic	172
5.3	Specialized inflections	184
5.3.1	<i>Be</i> -participle	188
6	The auxiliary: <i>have/be</i> alternations in the perfect	196
6.1	Evidence	196
6.1.1	Theoretical background	203
6.2	Auxiliary selection independent of transitivity/voice	208
6.2.1	Auxiliary selection according to person	209
6.3	Splits according to transitivity/voice	216
6.3.1	Auxiliary selection according to voice	216
6.3.2	Auxiliary selection according to transitivity	218
6.3.3	Irreversibility	222
6.4	Finer parametrization	223
6.4.1	Interactions between auxiliary selection according to transitivity/voice and according to person	224
6.4.2	The third auxiliary	228
6.5	Some conclusions	233

7	The noun (phrase): agreement, case and definiteness in an Albanian variety	236
7.1	Theoretical and empirical background	237
	7.1.1 Nominal inflections in Albanian	239
	7.1.2 Generative approaches to case	244
7.2	Analysis of Albanian nominal inflections	246
	7.2.1 Consonantal inflections	250
	7.2.2 Vocalic case inflections and lack of inflections	255
	7.2.3 Prepositional contexts	259
	7.2.4 Summary	261
7.3	The Albanian noun phrase	262
	7.3.1 The genitive	262
	7.3.2 The adjective	266
	7.3.3 Adjectives as heads of the noun phrase	272
7.4	Concluding remarks	275
8	(Definite) denotation and case in Romance: history and variation	276
8.1	The Latin case system	277
8.2	Romance case systems: Romanian	286
8.3	Other Romance case systems – and alternative accounts	295
8.4	Loss of case in Romance: Romansh –s	302
8.5	Pronouns – and some conclusions	308
	<i>Notes</i>	312
	<i>References</i>	331
	<i>Index</i>	345

Tables

6.1	Distribution of <i>be</i> (E) and <i>have</i> (A) according to person in the present perfect (in Central and Southern Italian varieties)	<i>page</i> 212
6.2	Distribution of <i>be</i> , <i>have</i> and <i>syncretic forms</i> in the present perfect in Piedmontese and Lombard varieties	231
7.1	Distribution of nominal inflections in Albanian	244
7.2	Denotational properties of Albanian nominal inflections	262
8.1	Denotational properties of Latin nominal inflections	286

Acknowledgements

The research reported in this book has been financed largely through PRIN grants from the MURST/MIUR, namely *Per una cartografia strutturale delle configurazioni sintattiche: microvariazione nei dialetti italiani* (1997–1999), *La cartografia strutturale delle configurazioni sintattiche e le sue interfacce con la fonologia e la semantica. Parametri morfosintattici e fonosintattici* (1999–2001), *Categorie linguistiche: Categorie di flessione nominale e verbale (Accordo, Aspetto); Nome e Verbo* (2001–2003), *I sistemi linguistici ‘speciali’ (apprendimento, disturbi) e la variazione tra i sistemi linguistici ‘normali’*. *Categorie funzionali del nome e del verbo* (2003–2005), *Strutture ricorsive in sintassi, morfologia e fonologia. Studi sulle varietà romanze, slave e albanesi* (2005–2007), *Morfosintassi e lessico: Categorie della flessione nominale e verbale* (2007–2009).

Special thanks go to all our informants, both Romance and Albanian, though space limitations prevent us from mentioning all of them here. Our debt to the friends and colleagues whose work inspired ours should be obvious from the references. However, we take this opportunity to thank Neil Smith, as a (former) general editor of the series, for helping our project along.

Introduction: grammatical categories and the biolinguistic perspective

According to Chomsky (2000b: 119), ‘the human language faculty and the (I-)languages that are manifestations of it qualify as natural objects’. This approach – which ‘regards the language faculty as an “organ of the body”’ – has been labelled the ‘biolinguistic perspective’ by Chomsky (2005: 1). Hauser, Chomsky and Fitch (2002: 1570) base their discussion of the key biological question of evolution on the ‘biologically and individually grounded’ use of the term language ‘to refer to an internal component of the mind/brain (sometimes called “internal language” or “I-language”)’. They distinguish two conceptions of the faculty of language, one broader (FLB) and one narrower (FLN):

FLB includes FLN combined with at least two other organism-internal systems, which we call ‘sensory-motor’ and ‘conceptual-intentional’ ... A key component of FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system and into the conceptual-intentional interface by the (formal) semantics system ... Most, if not all, of FLB is based on mechanisms shared with nonhuman animals ... FLN – the computational mechanism of recursion – is recently evolved and unique to our species. (Hauser, Chomsky and Fitch 2002: 1571)

The conception of the language faculty and of (I-)languages as ‘natural’, ‘biologically grounded’ objects corresponds to specific theories concerning their internal articulation:

the I-language consists of a computational procedure and a lexicon. The lexicon is a collection of items, each a complex of properties (called ‘features’) ... The computational procedure maps an array of lexical choices into a pair of symbolic objects, phonetic form and LF [logical form] ... The elements of these symbolic objects can be called ‘phonetic’ and ‘semantic’ features, respectively, but we should bear in mind that all of this is pure syntax and completely internalist. (Chomsky 2000b: 120)

The internal articulation of the FLN is crucial to the biolinguistic programme, no less than its applications to domains such as language evolution, genetics

2 *The biolinguistic perspective*

and neurology. Here we address some points concerning this; specifically, we concentrate on the issue of language variation, starting with the idea that ‘the diversity and complexity can be no more than superficial appearance ... the search for explanatory adequacy requires that language structure must be invariant’ (Chomsky 2000b: 7), and ‘There is a reason to believe that the computational component is invariant, virtually ... language variation appears to reside in the lexicon’ (Chomsky 2000b: 120).

From this perspective, a central aim of our work is to provide empirical support for what we may call the lexical parametrization hypothesis (Manzini and Wexler 1987), and thus to make more precise the sense in which it holds. Without a doubt ‘one aspect is “Saussurean arbitrariness”, the arbitrary links between concepts and sounds ... However, the possible sounds are narrowly constrained, and the concepts may be virtually fixed’ (Chomsky 2000b: 120). In the present study, we address the issue of how the linguistically relevant conceptual space yields different (I-)languages beyond the obvious aspect of ‘Saussurean arbitrariness’.

Before proceeding to the empirical core of the argument, we briefly introduce some of the conceptual underpinnings of the framework we adopt, beginning with the thesis that language ‘is a system that is, as far as we know, essentially uniform. Nobody has found any genetic differences ... since its emergence there has not been any significant evolution. It has stayed that way’ (Chomsky 2002: 147). This view is shared by much current work on human cognitive and linguistic evolution (Lieberman 1991; Jackendoff 2002). The conclusion holds both for living languages and for ancient ones (whether documented and no longer spoken or merely reconstructed); as argued by Labov (1994), the same mechanisms of (surface) variation and change affect all of them. To take a comparative typological perspective:

no evidence of anything like speciation has been found ... Languages from typologically very different areas have the same latent structural potential ... this survey has uncovered no evidence that human language in general has changed since the earliest stage recoverable by the method used here. There is simply diversity, distributed geographically. (Nichols 1992: 227)

As for this geographically distributed diversity:

a residual zone or a set of residual zones will contain a good deal of the world’s possible linguistic diversity in microcosm, and both the existence of internal diversity and its actual profile are stable and obviously very natural situations. Diversity of a particular kind may even be regarded as the state to which a group of languages will naturally revert if left undisturbed ... Spread zones, in contrast, are typically highly divergent from one another, but each is internally quite homogeneous ... Just which language spreads in a

spread zone is a matter of historical accident, and this historical accident can distort the statistical distribution of linguistic types in an area. (Nichols 1992: 23)

The set of languages considered in this work presents the kind of variation that we expect in natural languages in the absence of external constraints. Because of the political and cultural factors which, for centuries, have kept the Italian peninsula in conditions of great administrative and social fragmentation, dialectal differentiation in Italy has been preserved for longer (i.e. up to the present day) than in other areas of Western Europe, including Romance-speaking ones. Thus Italian varieties provide a rich and articulated picture of language variation that contrasts with that of other intensively studied varieties such as those of English. The view we take is that it is linguistic situations such as those in Britain, for example, that represent a somewhat misleading picture of variation, reflecting not only the internal shaping forces of language development, but also external mechanisms of social and political standardization. The variation seen in Albanian, including the major Gheg vs. Tosk divide in mainland Albania, and Arbëresh varieties of Southern Italy, has the same general character as that observed in Romance varieties. In the internalist (i.e. 'biologically, individually grounded') perspective that we adopt, variation between two or more varieties (linguistic communities) is in fact not qualitatively different from variation within the same variety (community), or even within the production of a single speaker. For example, to the extent that a speaker alternates between stylistic levels according to the situation of use, s/he will have a 'bilingual' competence of sorts – which, given the lexical parametrization hypothesis adopted here, can be accounted for as the co-existence of different lexicons with a single computational component (MacSwan 2000).

Suppose, then, that the lexicon is the locus of linguistic variation – in the form of a uniform (i.e. invariant) computational component, and of an invariant repertory of interface primitives, both phonological and conceptual. Non-trivial questions arise at this point: how can the lexicon vary on the basis of a universal inventory of properties (or 'features'), and why does that variation in the lexicon result in variation in order, agreement, selection, and other relations that are computationally determined? These questions are amply debated in current linguistic theory. Our empirical discussion aims to support certain positions emerging from the debate, as opposed to others which are in principle equally possible.

In particular, the answer to the preceding questions is mediated for various scholars by the notion that there is a fundamental distinction between functional and non-functional elements. Thus, within the Distributed Morphology framework, Embick (2000:187) assumes a 'distinction between the *functional* and *lexical* vocabularies of a language ... functional categories merely

instantiate sets of abstract syntacticosemantic features’, on which the derivational component operates. The actual phonological terminals corresponding to these abstract categories are inserted only after a level of morphological structure, where readjustment rules apply (Late Insertion). It is evident that the overall architecture of the grammar implied by this model is considerably more complex than one in which ‘the formal role of lexical items is not that they are “inserted” into syntactic derivations, but rather that they establish the correspondence of certain syntactic constituents with phonological and conceptual structures’ (Jackendoff 2002: 131).

Kayne’s (2006, 2008a) parametrization model, while avoiding recourse to Late Insertion, is close to Distributed Morphology in assuming that functional items correspond to a universal lexicon of sorts. Lexical and hence grammatical differences depend on whether the elements of this functional lexicon are overtly realized or ‘silent’. Interestingly, for Kayne (2006), even variation in the substantive lexicon can be reduced to variation in functional structure in the sense just defined, as can be seen in his construal of *shallow* as ‘LITTLE deep’, that is, essentially as the specialized lexicalization of *deep* in the context of the silent functional category ‘little’.

Manzini and Savoia (2005, 2007, 2008a) pursue a model under which, again, there is a unified conception of lexical variation – however, this is of the type traditionally associated with the substantive lexicon: there is a conceptual and grammatical space to be lexicalized and variation results from the distinct partitioning of that space. There is no fixed functional lexicon which varies along the axis of overt vs. covert realization – so-called functional space is just like all other conceptual space, and all lexical entries are overt. Thus, the distinction between functional (i.e. grammatical) contents and conceptual ones is an external one; as such it may very well be useless, and at worst it may obscure the real underlying linguistic generalizations.

Our conception of variation within the so-called functional lexicon is consistent with current conclusions regarding the conceptual space and the different ways in which it surfaces in natural languages. Fodor (1983) and Jackendoff (1994), among others, develop the Chomskyan theme that concepts, like other aspects of language, must have an innate basis – largely because of the poverty of stimulus argument. It has already been observed by Lenneberg (1967) that lexical items are the overt marks of a categorization process through which human beings carve out an ontological system from the perceptual continuum of the external world. This process of categorization is of course only indirectly connected with the objects of the external world. Jackendoff (1994: 195) notes that the lexical forms employed to express spatial location and motion (e.g. *The messenger is in Istanbul; The messenger went from Paris to Istanbul;*

The gang **kept** the messenger **in** Istanbul) typically also express possession (e.g. *The money is Fred's*; *The inheritance finally went to Fred*; *Fred kept the money*), the ascription of properties (e.g. *The light is red*; *The light went from green to red*; *The cop kept the light red*), etc.

This suggests that thought has a set of precise underlying patterns that are applied to pretty much any semantic field we can think about. Such an underlying 'grain' to thought is just the kind of thing we should expect as part of the Universal Grammar of concepts; it's the basic machinery that permits complex thought to be formulated at all. (Jackendoff 1994: 197)

Dehaene, Izard, Pica and Spelke (2006) study geometrical concepts in an isolated group of Amazonian people whose language, Mundurukú, 'has few words dedicated to arithmetical, geometrical, or spatial concepts'. They conclude that

geometrical knowledge arises in humans independently of instruction, experience with maps or measurement devices, or mastery of a *sophisticated geometrical language* ... There is little doubt that geometrical knowledge can be substantially enriched by cultural inventions such as maps, mathematical tools, or *the geometrical terms of language* ... however, the spontaneous understanding of geometrical concepts and maps by this remote human community provides evidence that core geometrical knowledge, like basic arithmetic is a universal constituent of the human mind. (Dehaene, Izard, Pica and Spelke 2006: 385, our italics)

In a similar vein, Hespous and Spelke (2004) study the acquisition of the conceptual distinction between 'tight' and 'loose' fit of one object to another in English-speaking children, which is not lexicalized in English, though it is in other languages like Korean. Their conclusion is that 'like adult Korean speakers but unlike adult English speakers, these infants detected this distinction ... Language learning therefore seems to develop by linking linguistic forms to universal, pre-existing representations of sound and meaning' (Hespous and Spelke 2004: 453).

In short, the building blocks that are combined to make up the potentially infinite variety of human lexicons are innate. The lexicons of different languages are formed on this universal basis, covering slightly different extensions of it and in slightly different ways. The view we advocate here is simply that ways of representing the event, such as transitivity or voice (chapters 5–6), ways of connecting arguments to predicates (or to one another), such as cases (chapters 7–8), and more, are to be thought of as part of this general system. There is no separate functional lexicon – and no separate way of accounting for its variation. We started with the general Chomskyan biolinguistic, or internalist, picture of language, and of its basic components, both broadly and narrowly construed. Variation is crucial to establishing this model for the obvious reason that the uniformity thesis, as laid out above, requires a suitably

restrictive account of observed cross-linguistic differences. But, even more fundamentally, the lexical parametrization hypothesis that we adopt means that questions of variation will inevitably bear on the form of the lexicon, as one of the crucial components of the I-language.

The other main component of the I-language is ‘the computational procedure’, which ‘maps an array of lexical choices into a pair of symbolic objects, phonetic form and LF’ (Chomsky 2000b, quoted above). As for the latter, Culicover and Jackendoff (2005: 6) aptly characterize a particularly popular conception of the relation of LF to the syntax (i.e. the computation) as ‘Interface Uniformity’, which holds that ‘the syntax-semantics interface is maximally simple, in that meaning maps transparently into syntactic structure; and it is maximally uniform, so that the same meaning always maps onto the same syntactic structure’. This bias inherent in much current theorizing provides a standardized way of encoding the data, but does not appear to have any strong empirical motivation; nor is the encoding it provides a particularly elegant or transparent one. Conceptually it corresponds to a picture where syntax ‘includes’ interpretation, in the sense that all relevant semantic information finds itself translated into syntactic structure. In contrast, we agree with Culicover and Jackendoff (2006: 416) on the idea that interpretation is ‘the product of an autonomous combinatorial capacity independent of and richer than syntax’, ‘largely coextensive with thought’, which syntax simply restricts in crucial ways.

Linguistic meanings are merely an input to general inferential processes; the linguistic categorization of the conceptual space encoded by lexical items does not correspond to ‘meaning’ itself but rather to a restriction of the inferential processes producing it. Sperber and Wilson (1986: 174) provide a particularly compelling discussion of the point that linguistic expressions only denote because of their inferential associations: ‘Linguistically encoded semantic representations are abstract mental structures which must be inferentially enriched’. In such a model, the well-known indeterminacy of linguistic meanings becomes a key property of successful communication:

A linguistic device does not have as its direct proper function to make its encoded meaning part of the meaning of the utterances in which it occurs. It has, rather, as its direct proper function to indicate a component of the speaker’s meaning that is best evoked by activating the encoded meaning of the linguistic device. It performs this direct function through each token of the device performing the derived proper function of indicating a contextually relevant meaning. (Orizzi and Sperber 2000: 160)

Note that we disagree with Culicover and Jackendoff (2005) on the model of syntax to be adopted. Our analysis depends on a representational version of minimalism, roughly in the sense of Brody (2003). Crucially, the LF primitives

we employ are independently available within a minimalist grammar as defined by Chomsky (1995), and in this sense the approach we take is compatible with Chomsky's model. In fact, we would argue that our views on lexical variation and on interpretation are the simplest construal of Chomsky's (2000b) proposals, as summarized above – much simpler than other current approaches, and in this sense closer to the core of minimalism and of the biolinguistic programme.

Therefore, any theory maintaining a functional/lexical divide must define the boundary between the two – which is a far from trivial task. The domain of spatial relations and of events involving them is a case in point. Spatial relations are covered by prepositions (or particles in their intransitive use), among other items. In particular, prepositions/particles can combine with elementary verbs to lexicalize events with a spatial component; for instance, English has *put down* (*the book*), Northern regional Italian has *mettere giù* (*il libro*). At the same time, Tuscan and literary Italian has a verb *posare* 'put down', and the examples could be multiplied (*go in* and *enter* in English, etc.). Particles in Germanic languages (but also in Romance, for instance in Northern Italian varieties) also allow for aspectual interpretations. If, on the basis of these, of the role they play in case systems, etc., we treat prepositions/particles as part of the functional lexicon, what should we infer about spatial primitives? Are they functional? If so, how is their relation to *posare*, *enter*, etc. (i.e. canonical lexical verbs) expressed?

As mentioned above, the answer envisaged by authors such as Kayne (2006) is that apparent variation in the substantive lexicon reduces to variation in the pronunciation of functional categories; hence the substrings lexicalized by what would traditionally be thought of as lexical categories consist in reality of a number of functional specifications – which may surface in some languages and not in others, or surface to different extents in different languages. In this way, the functional lexicon effectively spreads over considerable portions of the substantive lexicon; taking this to the extreme, one may want to say that lexical categories are but an epiphenomenon of abstract functional structure.

Since the proposal we are putting forward is that lexicons are merely ways of partitioning an abstract categorial space, we are in a way suggesting theories close to those we are taking issue with. At the same time, we consider it significant that we take the step of calling the lexical/functional divide into question, while they typically don't. To begin with, the different approaches make different empirical predictions in the data domains they both address. Thus, we have specifically referred to Kayne (2006, 2008a, 2009) and Distributed Morphology, since we can directly compare our respective approaches with regard to such domains as fine variation in clitic structures, where we believe our model to be preferable on grounds of descriptive as well as explanatory adequacy (Manzini and Savoia 2009b, 2010).

The lexical/functional issue seems to us particularly noteworthy, because at heart it concerns the distinction between the narrow and broad language faculty (FLN and FLB). Let us assume that there is a universal inventory of concepts, and that the lexicon represents a way of realizing it. In theories in which there are in fact two inventories, one for functional categories and one for non-functional ones, it seems to us that the functional and non-functional lexicons are implicitly or explicitly apportioned to the language faculty narrowly construed and broadly construed, respectively. The reduction of the divide that we are proposing has implications not only for the more technical aspects of the theory of grammar, but also opens up the possibility that the universal conceptual repertory which is partitioned by language-particular lexicons is part of the broadly construed language faculty in its entirety. In fact, we see no reason why the grammatically relevant categories investigated here should not constitute categorizations in a domain of general cognition. In other words, what we are saying is that the existence of a functional lexicon associated with the FLN is not a matter of logical or factual necessity – and as such it should be open to scrutiny.

Given the position that we tentatively take on the matter – namely that eliminating the divide does not imply any empirical problem, and on the contrary allows for a certain simplification of the architecture of language – we may wonder why such a distinction is so prominent in linguistics. Neuropsychological literature provides much evidence, based both on recent brain imaging techniques and on more traditional language disorders and acquisition studies, that different brain areas are implicated by different conceptual clusters. The prediction is that

manipulable objects such as tools are strongly linked to motor behaviour and therefore their representational networks should comprise a significant amount of neurons in motor contexts. Animals, which are most of the time (visually) perceived rather than manipulated, should be represented by networks that partly reside in the visual cortex. (Bastiaansen *et al.* 2008)

Conversely, ‘assemblies representing function words remain limited to the perisylvian cortex and strongly left-lateralized in typical right-handers’ (Pulvermüller 1999: 260–1). This appears to underlie, in particular, the differential treatment of different sublexicons by aphasic patients (anomics, agrammatics, etc.). Given such results, it does not seem to us to be necessary to draw the conclusion that there is a functional lexicon associated with the computational system of natural language and distinguished on these grounds from a contentive lexicon. Another possibility is that

there is a continuum of meaning complexity between the ‘simple’ concrete content words that have clearly defined entities they can refer to ... more abstract items that may or may not be used to refer to objects and actions

and function words ... According to the present proposal, the important criterion is the strength of the correlation between the occurrences of a given word form and a class of non-linguistic stimuli or actions. (Pulvermüller 1999: 261)

In other words, it is not so much the functional lexicon that has a special status within the architecture of the mind-brain, but rather certain concrete contents as opposed to more abstract ones.

Once freed from the burden of highly articulated inventories and hierarchies of functional categories, we can entertain a simpler syntax, much in the sense of Culicover and Jackendoff (2005). As already mentioned, on the other hand, we do not believe that levels of representations of the type proposed by Culicover and Jackendoff (2005), including rich notions such as grammatical functions, linking rules etc., are required by such a simpler syntax. Rather, the grammar implemented here is a representational version of current minimalist theories (cf. Brody 2003).

The relation of the syntax, and more precisely its LF component, to interpretation, as outlined above, is crucial in our view to understanding the role of language variation in the overall economy of the faculty of language. If our construal of syntax and its relation to interpretation is correct, the syntax restricts interpretation, but does not 'contain' it (Culicover and Jackendoff 2006). Thus the boundary between syntax and interpretation is a loose one, allowing for a number of different matchings of syntactic form to (inferentially determined) meaning. The looseness of this relation seems to be an essential design feature of the faculty of language, in the sense that it permits the invariant constructs of syntax to cover changing meanings. Lexical items are at the core of language variation simply because they represent the core unit of this interface between syntax and interpretation. In this sense, variation is not an accidental property of the faculty of language, and neither are the characteristics of variation that we try to outline in this study. Rather, they pretty much represent a by-product of the general design of the language faculty.

The aspect of our work which provides the title for this book ('grammatical categories') has to do with the redefinition of the grammatically relevant classes (i.e. the 'categories') of natural language. In general, we take it that the lexicons of natural languages are learnable in that lexical entries individuate natural classes. We apply this logic in particular to Romance complementizers which have the same form as *wh*-items (Italian *che* and the like) and to Romance sentential negations which have the same form as negative polarity arguments, in particular 'nothing' (Piedmontese *nen* etc.). In both cases we conclude that lexical identity of form is not a matter of homophony but reveals the sharing of deeper categorizations. This calls into question, among

other things, the classical functional categories of C(OMP) (chapters 1–2) and NEG (chapters 3–4). Elsewhere in this book, we find no reason to entertain a functional category status for the so-called AUX(iliaries) *have* and *be*, which are argued just to be main verbs selecting a participial clause (chapter 6). In chapter 5 the cluster of meanings associated with Romance *si* and its Albanian counterpart *u* are reduced to a unified characterization which also holds of other morphological instantiations of middle-passive voice. Even syncretisms involving case morphology – and the functional category K(ASE) according to some (cf. Fillmore 1968; Giusti 1995), are analysed in chapters 7–8 as instances of ambiguous interpretation of the same underlying category, rather than as instances of default lexicalization. This, in turn, requires a revision of the categorizations provided by standard morphological feature systems.

It should be kept in mind that the functional structure that this book calls into question (COMP, NEG, AUX, K) is quite independent of recent cartographic proposals (see Cinque and Rizzi (2008) for an overview) which aim to provide a fine-grained picture of functional categories and the way in which they map to syntactic hierarchies. The result is an increase in the number of functional categories, yielding hierarchies of considerable complexity, which have been objected to on the grounds that they enrich the grammar by introducing a great number of new categories and orderings. Yet the same concern regarding the expressive power of the theory could be voiced for standard approaches to functional structure, since the creation of a new functional category or a new feature annotation of an existing category is not subject to any formal or substantive constraints.

In this book we propose a take on the problem which goes back to the very first models of exploded structures (Larson 1988), and even further to the very first approaches to ‘functional’ structure in generative grammar (Rosenbaum 1967 on complementation). We argue that structures are indeed atomized, in the sense that a wealth of differentiated head positions are projected under Merge. At the same time, our contention is that a considerable amount of this atomization (perhaps all) does not derive from the introduction of novel categories, but simply from the recursion of certain elementary, identical cells.

Thus, the complementizer (chapters 1–2) is not introduced as a specialized head C(OMP); rather, the clearly nominal nature of the complementizer in Romance languages (as in Germanic ones) suggests that the complementizer is the N complement of the matrix verb; in turn, this N takes the embedded sentence as its complement. This structure is as internally articulated as that of Rizzi (1997), but its internal articulation does not depend on a functional hierarchy. Rather, it depends on the recursion of ordinary nominal and sentential

embeddings. Similarly, negation (chapters 3–4) is one of the earliest functional categories proposed under an articulated view of phrase structure, dating back at least to Pollock (1989). Based on evidence from Romance varieties, we propose, however, that so-called negative adverbs and heads are negative polarity elements, and, even more radically, that they participate in the argumental structure of the verb, coinciding specifically with the individuation of the internal argument position. Thus, in Romance languages, there is neither evidence for a lexicalized negative operator nor for a functional position hosting it.

The final case study to be introduced here concerns the internal structure of nouns and noun phrases (chapters 7–8). Following an established trend in generative grammar, we argue for the conclusion that noun phrases (as well as adjective phrases) have the same internal organization as sentences. From this perspective, we take up the classical proposal of Higginbotham (1985) that the D(eterminer) saturates the obligatory (internal) argument of the nominal predicate; in this sense, D properties yet again represent an instantiation not of functional structure, but of predicate–argument structure. Case, in turn, is not construed as a (functional) primitive of grammar, but rather as a label covering much more elementary properties, relating again to the saturation of predicate–argument structures.

Throughout the discussion, the emphasis is very much on empirical evidence. We repeatedly argue that our model not only fares better with respect to fairly reasonable simplicity metrics, but also that it has descriptive advantages. In fact, and quite strikingly in our opinion, less powerful theories are better suited to capturing complex (micro)variation data of the type we consider than theories potentially capable of greater descriptive power.

In particular, we subscribe to the simplicity argument in favour of representational grammars advanced by Brody (2003). This implies abandoning derivations, including the notions of a cycle (phases) and an asymmetric search space (feature checking). What we retain is representational relations: chains, agreement, etc. Simplicity is paramount, to the extent that existing empirical evidence does not provide any support for the more complex grammar. In particular, complex data concerning agreement (and variation in agreement patterns) are accounted for in Manzini and Savoia (2005, 2007, 2008a) by abandoning phi-feature checking in favour of identity (or better, compatibility) of referential properties; uninterpretable and unvalued features are also eliminated under this approach. At no point is there any evidence that a derivational approach would have empirical advantages – on the contrary, the complexity of the variation effectively requires the simpler representational approach. In general, representational grammars are simpler than derivational ones in that the latter postulate

purely computational processes whose results are LF-relevant and hence redundant with LF constructs. Movement and the LF-relevant notion of chain are the obvious cases, but this also holds of the computational operation of agreement and its LF reflexes relevant for coreference etc. By contrast, the representational model views LF-relevant relations as determined directly by the interpretive calculus at the LF interface (chains by the theta-calculus, and so on).

This adoption of a representational model goes hand in hand with the present take on functional structure. In the minimalist grammar of Chomsky (1995), functional categories, or features, bear the burden of computation, since they crucially enter into feature checking. To the extent that functional categories are reduced to ordinary predicates and arguments, functional feature checking is also replaced by ordinary selectional restrictions. In turn, empty functional heads and/or Specs, needed only for the checking of functional features, can be eliminated. As for the highly articulated hierarchies proposed in cartographic research, if functional and lexical properties of lexical entries are not formally different, as argued here, the most natural and economical assumption is that their ordering is restricted entirely by interpretive principles, including closure requirements, scope, etc.

Needless to say, in the absence of any theoretical distinction between functional and lexical elements, we do not expect that they will be inserted at two different points in the derivation, namely in syntactic and morphological structure respectively (as in Distributed Morphology). On the contrary, we predict that all syntactic structure will be projected directly from lexical entries, independently of their properties. In other words, we propose a view in which morphological-level structures and relations are entirely unified with syntax. Lexical entries, in turn, are entirely characterized in terms of positive properties.

Another theme of general significance is the question of how the present approach to functional structure relates to the issue of ‘grammaticalization’. In the terms of, say, Roberts and Roussou (2003), grammaticalization is essentially the reanalysis of a lexical category as a functional category, often with the result that a lexical entry can have both a lexical and a functional construal. In such cases we typically propose that there is a single lexical item with a unified (lexical, not functional) characterization; the *che* ‘that’ complementizer/*wh*-phrase (chapters 1–2) is a case in point. It is evident that, to the extent that this latter treatment can be generalized, the distinction between functional and lexical dissolves; if so, the problem of why exactly lexical categories would turn into functional ones (i.e. the problem of ‘grammaticalization’) simply does not arise.

1 *The structure and interpretation of (Romance) complementizers*

In this chapter we start from the observation that in Romance languages, complementizers are nominal, belonging to the same argumental series as *wh*-phrases, although current theories treat them as functional projections of verbs, filling the same positions as verbs do. We argue that the Romance *che*-type complementizer is not a functional category of the verb, but rather a nominal head, which satisfies an argument slot of the matrix verb and which takes the embedded sentence as its complement (section 1.1). Both as a *wh*-phrase and as a complementizer, Italian *che* introduces a variable. If it introduces an individual variable, the *wh*-phrase reading arises; if it introduces a propositional variable (ranging over situations or possible worlds) it is read as a so-called complementizer.

In section 1.2, we argue that our proposal is compatible with fine distributional evidence relating to the left periphery of the sentence, in particular with the fact that elements such as topics or foci can occur both below the complementizer and above it. We capture this distribution by allowing them to occur at the left periphery of the embedded sentence (below the complementizer) or at the left periphery of the complementizer phrase (above the complementizer). This structural hypothesis predicts that occurrences of the relevant material below and above the complementizer can combine – a fact that requires quite complex functional hierarchies in alternative accounts. It should be kept in mind that, insofar as this chapter proposes an alternative to the generally adopted view that the complementizer is a functional projection of the sentence, it is not aimed specifically at so-called cartographic models. These, however, are discussed in some detail in section 1.3 (especially Rizzi 1997, 2001, 2004), because they explicitly consider the same type of data as we do.

In section 1.2 we also briefly consider possible general objections to our proposal, for instance concerning the distinction between complementizer phrases and conventional noun phrases: how is this distinction made for the purposes of selection? Extraction facts are also relevant: how are sentences

introduced by complementizers different from complex NPs? For the purposes of selection, complementizer phrases can be distinguished from conventional noun phrases in that they introduce a propositional and an individual variable respectively. As for extraction, the question why complementizer phrases, in contrast to conventional noun phrases, are not islands is the same as the question why complementizer phrases have a left periphery that is an ‘escape hatch’ (Chomsky 1973), and conventional noun phrases do not. This remains very much an irreducible primitive of all generative approaches (up to Chomsky 1995, 2008), since it is not the case that noun phrases lack a left periphery altogether (Szabolcsi 1994). We construe this primitive not as a structural, but as an interpretive one, contrasting propositional denotations with individual denotations.

A different question concerns the fact that, although Romance complementizers belong to the *wh*-series, they typically have a non-interrogative interpretation. We return to this question in [chapter 2](#), where we show that this property is parametrized. In particular, we illustrate Romance systems where the ‘that’ complementizer – or an element of the *wh*-series – lexicalizes the ‘if’ (interrogative/ hypothetical) complementizer as well.

In general, the argument in favour of the present theory is based on explanatory adequacy. First, it simplifies the lexicon, allowing for a unified lexical entry for elements like Italian *che*, the only possible alternative being homophony. It also solves the mystery of why C would host such disparate categories as verbs and complementizers, reserving what we might continue to call C for verbs. Perhaps most interestingly, it reduces what would otherwise be potentially complex functional hierarchies to the recursion of simple predicate–argument structure (i.e. the complementizer is an argument of the matrix predicate, taking the embedded sentence as its argument). Furthermore, the burden of proof is on alternative theories to prove that they can account in an explanatory way for the fine variation we observe in complementizer systems, a matter to which we return in [chapter 2](#).

1.1 Romance complementizers are nominal and head their own noun phrase

Current theories hold that complementizers are functional projections of the verb, that is, their position is essentially one which could in other circumstances be filled by a verb. This is true independently of whether there is a set of C positions (Rizzi 1997), or a single C position with a rich specifier structure (Chomsky 1995). But if complementizers fit into the same type of position as

verbs, it is unclear why they so clearly coincide with functional projections of the noun: for instance with demonstratives in Germanic, or with *wh*-elements in Romance. Even if we were to accept that the relation of complementizers to the *wh*-system of Romance languages, to the demonstrative system of English etc. is a purely historical one, the question would arise of how nominal, quantificational elements could come to fit verbal specifications. In other words, the form of the problem would change, but not its substance.

Consider for instance Italian, in which *che*, like English *that*, introduces finite declaratives, as in (1); since Kayne (1976) this has been identified with the relative clause introducer in (2). The same element also introduces interrogatives with the meaning of ‘what’ as in (3). Furthermore, *che* can appear as the *wh*-determiner of complex interrogative NPs as in (4).

- (1) So che fai questo
 I.know that you.do this
 ‘I know that you do this’
- (2) Il lavoro che fai è noto
 the work that you.do is known
 ‘The work you do is well-known’
- (3) Che fai?
 what you.do?
 ‘What are you doing?’
- (4) Che lavoro fai?
 which job you.do
 ‘Which job do you do?’

One possibility that we can reject is that Italian *che* simply has two lexical entries, one of which corresponds to the ‘that’ complementizer and the other to the ‘what’ *wh*-element. This solution does not have any explanatory value, given that the pattern that it describes is not an accidental coincidence observed in one or even a few languages, but a systematic phenomenon in Romance, as can be gleaned from the data to follow. What is more, Caponigro and Polinsky (2008) find the same formal identity between the *wh*-system and the complementizer system in a completely unrelated language, Adyghe (a NW Caucasian, Abkhazo-Adyghean language), showing that in Adyghe the same syntactic structure can be mapped to four different meanings, namely relative clause, complement clause, *wh*-interrogative and yes–no interrogative.

Uncontroversially, in sentences like (3) *che* heads its own noun phrase, which in turn fills a position in the C field of the sentence, as in (5a). From an interpretive point of view, *che* introduces a variable corresponding to the internal argument of the predicate *fare* ‘to do’, as in (5b).

- (5) a. [che] [C [I fai
 b. che x, fai x

In examples of the type in (4), the interrogative noun phrase *che lavoro* ‘what job’ is associated with a structure of the type in (6a), where *che* represents a quantificational position Q within the noun phrase. Assuming that the structure of the noun phrase is parallel to that of the sentence, the inflected head noun is positioned in I – to parallel the position of the inflected verb within the sentence. From an interpretive point of view, *che* again introduces a variable restricted by the predicate *lavoro* ‘job’, as in (6b).

- (6) a. [Q che [I lavoro]] [C [I fai
 b. che x: x lavoro, fai x

Apart from the fact that both complementizer and *wh*-phrase take the morphophonological form *che*, from a syntactic point of view the complementizer *che* in (1) and the *wh*-phrase *che* in (3) both occupy a position at the left periphery of the sentence. Since under current theories, complementizer *che* fills precisely the kind of position that we questioned at the outset, namely C, in (7a) it is assigned to an unnamed position above C – the latter independently known to be a possible position of the verb in the sentence. From an interpretive point of view, complementizer *che* could be an operator introducing a variable, again like *wh*-phrase *che* – the main difference between them being the nature of the variable. This ranges over individuals for the *wh*-phrase *che* in (5)–(6), while for complementizer *che* it ranges over situations/possible worlds, as in (7b).¹

- (7) a. [che [C [I fai questo
 b. che x: x fai questo

The fact that the operator-variable structure is interpreted as a question in (5)–(6), but not in (7), need not stand in the way of their unification, as we know that *wh*-phrase *che* can also have non-interrogative interpretations (very much like its English counterpart *what*). Thus *che* can be the head of a free relative, as in (8), or an exclamative, as in (9). The most natural interpretation of this range of data is that question, declarative (relative) and exclamative values are not intrinsic to the *wh*-phrase, but rather contextually determined.²

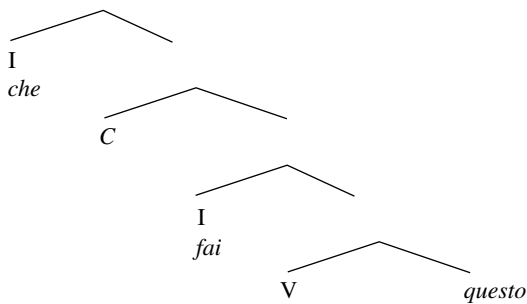
- (8) Fai che ti pare
 do what to.you pleases
 ‘Do what you like’

- (9) Che non farei!
 what not I.would.do
 ‘What I wouldn’t do!’

Given the discussion so far, if complementizer *che* appeared in the appropriate context, there is no reason why it shouldn't assume the interrogative value of, say, English *if* or *whether*. Now, in well-known Romance languages like Italian, 'if' has a specialized lexicalization, *se* in Italian, in which the hypothetical and interrogative values overlap. In [chapter 2](#), however, we will show that this state of affairs does not necessarily hold – in other words, that there are languages in which the lexicalizations of 'that', 'if' and 'wh-' overlap, as we expect. Similarly, recall that the Adyghe morphology discussed by Caponigro and Polinsky (2008) is ambiguous between a 'yes–no interrogative' and a 'complement clause' interpretation (among others).

Let us assume, then, that complementizer *che* is to be identified with *wh-*phrase *che* (as the morphology but also the interpretation suggest) and that, as anticipated in the discussion of (7a), it cannot be hosted in C, since C is a verbal position, while *che* is nominal. An alternative structure is suggested by what is perhaps the earliest approach to finite complementation in generative grammar (Rosenbaum 1967) as well as by analyses of special subsets of complement sentences, notably factives (Kiparsky and Kiparsky 1970) and unselected questions (Adger and Quer 2001). These have in common the postulation of a nominal structure of some sort for sentential complementation. What we propose here is that every sentence introduced by *che* has a nominal layer, represented by a noun phrase headed by *che* itself. Thus, in a sentence like (1) the verb heading the matrix sentence takes as its complement a noun phrase headed by the *che* complementizer, which in turn takes the embedded sentence as its complement, yielding a structure of the type in (10). Following the assumptions we have already introduced in (6) concerning the position of head nouns in noun phrases, *che* is in I.³

(10)



It should also be noted that the theoretical literature does contain treatments in which sentential introducers are removed from the C projection(s) of the embedded sentence and are made to depend directly on the main sentential tree. Notably, Kayne (1994), in analysing the infinitival introducers of Romance languages, such as Italian *di* ‘of’ etc., treats them not as complementizers, but as functional projections of the main verb, triggering movement of the embedded sentence to their Spec position. What we propose here is different, in that we are rejecting entirely the idea that complementizers are functional heads (either of the embedded or of the main sentence) and we are arguing that they are true arguments of the main verb, in turn taking the embedded sentence as their argument. For reasons of space we will not be able to deal with *di*-type introducers here. Manzini (1982) and Manzini and Savoia (2005) treat them as prepositions.

A more direct comparison can be established with the work of Arsenijevic (2009), according to whom ‘the variable denoted by a *wh*-element gets bound by a question operator in questions or by an appropriate head in relatives. Its presence in F[inite] C[omplement] C[lausel]s in a significant number of languages signals that the denotation of these clauses involves a variable’. This conclusion converges with those in the text, yet the syntax proposed by Arsenijevic (2009) for English markedly differs from ours. Thus, *that*, which semantically introduces a lambda operator, syntactically occupies the conventional C position. The variable over which the lambda operator abstracts corresponds to the Spec of a Force head generated immediately under the C head. This position is occupied either by a nominal expression with Force content such as *claim* or by its abstract incorporated counterpart for verbs such as *to claim* (analysed as *make claim*). In these terms the overall structure assigned to *claim that John kissed Mary* is a relativization headed by *claim*: [_N claim [_C that [_{Force} ~~claim~~ [John kissed Mary]]]].

Similarly, Kayne (2010) revises his (1976) idea that the relative *que* of French is really the *que* complementizer, by proposing instead that ‘*that* isn’t [a complementizer]. The *that* that introduces sentential complements is really a relative pronoun ... The claim that English sentential *that* is a relative pronoun must be taken to extend ... for example, to Italian *che*, to French *que* ... From the present perspective *that* is not a Force in Rizzi’s (1997) sense nor the head of a CP phase in Chomsky’s (2001) sense’. The actual implementation that Kayne provides for these ideas differs markedly from ours, while bearing a considerable resemblance to Arsenijevic’s. In particular, for Kayne ‘factive sentences ... have a deleted or silent FACT. If so ... factives too must involve relative clause structures.’ Assume, then, that sentential complements and

sentential subjects are always accompanied by a head noun, even if they are not factive ... Either it will be raised from within the relative in a way largely parallel to what happens with overt fact ... or a silent noun will be so raised’.

We have insisted more than once on the empirical reasons, but also on the general theoretical problems that lead us to reject the silent categories approach of Kayne, for instance in the domain of clitic phenomena (Manzini and Savoia 2008a, 2009a, 2010; Savoia and Manzini 2010). These effectively apply to Arsenijevic (2009) as well. On the other hand, it seems to us that the extra assumptions concerning the (silent) Force phrase of Arsenijevic, or equivalently the silent head nouns of Kayne, are unnecessary, since both the semantics for propositional embedding and the identity of so-called complementizers and *wh*-phrases follow from the simpler syntax proposed here. We will return to this comparison in chapter 2.

In the next section, we argue that the structure in (10) is compatible with the distributional evidence concerning the so-called left periphery of Romance languages – in fact, it provides a particularly economical way of dealing with it.

1.2 Structure of the complementizer phrase

Given a structure like (10), we predict that material related to the quantificational and informational structure of the sentence (*wh*-elements, topic, focus, etc.) should be hosted by the left periphery of the embedded sentence and hence should follow the complementizer. Indeed the presence of such material under the complementizer is familiar from English as well as from Romance languages. In (11) we provide some examples from Italian, where (11a) illustrates a topic and (11b) a focus in the left periphery of the embedded sentence.

- (11) a. So che questo non l’hanno preso
 I.know that this not it they.have taken
 ‘I know that they haven’t taken this’
- b. So che questo hanno preso (non quello)
 I.know that this they.have taken (not that)
 ‘I know that they have taken this one, not that one’

As expected, *wh*-phrases can also appear under the complementizer. In fact, in our own judgement, this option is open in Italian, as in (12). In any event it is fully productive in Southern Italian varieties such as *Arena* in (13), where any *wh*-phrase can appear either in the left periphery of the matrix sentence – i.e. in its scope position, as in (13a’) and (13b’) – or in the left periphery of the embedded sentence, hence under the complementizer, as in (13a) and (13b).

(12) *Credi che con chi se la prenderanno*
 you.think that with whom MP it they.will.take
 'Who do you think they will take it out on?'

(13) *Arena* (Calabria)

- a. *ti kriði ka ðuvi ðormi*
 you think that where he.sleeps
 'Where do you think he sleeps?'
- a'. *duvi ti kriði ka ðormi*
 where you think that he.sleeps
 'Where do you think he sleeps?'
- b. *ti kriði ka pək'ki veni*
 you think that why he.comes
 'Why do you think he is coming?'
- b'. *pək'ki ti kriði ka veni*
 why you think that he.comes
 'Why do you think he is coming?'

In turn, the verb can invert with the subject in sentences embedded under the complementizer, yielding instances of embedded V2. In particular, among the (present-day) Romance languages, many Romansch and Ladin varieties, like *La Pli* and *Scuol* in (14) and (15) respectively, exhibit V2 in embedded sentences, hence under a *che*-type complementizer. On the standard assumption that V2 depends on the positioning of the verb in C, this means that the left periphery of sentences embedded under the complementizer, i.e. its C field, can host a verb as well.

(14) *La Pli de Mareo* (Alto Adige/South Tyrol)

- a. *i te diži ke le liber a-i lit*
 I you tell that the book have-I read
 'I am telling you that I have read the book'
- b. *i te diži ke endo'maŋ/ go'not/ magari veŋ-el*
 I you tell that tomorrow/ often/ may be comes-he
 'I am telling you that he is coming tomorrow/ often/ perhaps'

(15) *Scuol* (Grisons)

- i m an dit tʃa fərs dərmaʃ-t/ dər-m-al*
 they me have told that perhaps sleep-you/ sleeps-he
 'They told me that perhaps you are/ he is sleeping'

The data that have been presented are merely compatible with the structure in (10) and do not in themselves provide any argument in its favour. An adequate analysis of the same data is available within the articulated theory of the C field proposed by Rizzi (1997, 2001, 2004). In particular, the distribution of Italian *che* and other Romance *che*-like complementizers in (11)–(15)

corresponds to that predicted by Rizzi (1997) for the highest complementizer.⁴ The embedded V2 facts follow, as already proposed by Schwartz and Vikner (1996), if the verb occupies a different C position (in fact the lowest C position for Rizzi (1997)).⁵

By contrast, the existence of embedded V2 is problematic in frameworks like Chomsky's (1995), where only one C position is available. In such a framework, one possibility is to derive embedded V2 from the positioning of the verb in I (Santorini 1989). The clitic nature of the inverted subject in (14)–(15) tends to discount this possibility, on the assumption that subject clitics are associated with a high inflectional domain – hence if the verb were in I, subject clitics should precede, rather than follow, the verb. The only alternative that we can see to Rizzi's (1997) articulated view of the C field is precisely the one we are advocating here: namely, treating the complementizer as the head of an independent noun-phrase-like projection.

The discussion so far only addresses the conditions that make structures like (14)–(15) possible; a different question altogether is what makes them necessary. According to McCloskey (2004) there is a particular interpretive value associated with embedded V2, which is revealed by the selectional restrictions discussed in the literature on Germanic languages. Thus, the highest layer of an embedded V2 structure – i.e. the one filled by the complementizer – realizes illocutionary force, and allows for inversion of the verb in the lower layer. In languages/contexts that do not allow for embedded V2, a single layer is present, hosting the complementizer and excluding V2; this does not realize illocutionary force.

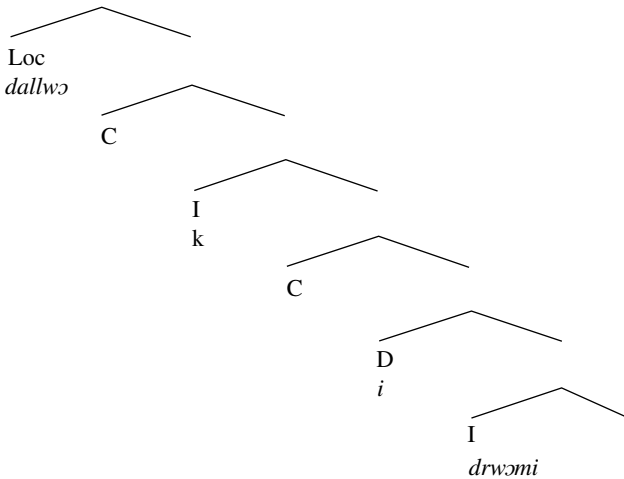
We are not aware of any data on Romance embedded V2 that would allow us to decide whether it obeys semantic restrictions. But suppose it does. The general schema of explanation proposed by McCloskey (2004) need not be tied to the double CP structure that he adopts; rather, it can be implemented, as far as we can see, by the structure that is at the heart of the present proposal. Thus, properties of the matrix sentence select a particular set of properties on the embedded complementizer – which in turn select for V2 in the embedded sentence. Languages/contexts without embedded V2 simply do not select for the relevant properties on the complementizer head, and so the V2 position of the verb is not selected in the embedded sentence.⁶

While in (12)–(13) we have seen some cases where the *che*-type complementizer precedes a *wh*-phrase, there are many Romance varieties in which the *che*-type complementizer follows a *wh*-phrase in both main and embedded interrogatives. This pattern is quite robust in Northern Italian varieties; in (16) we exemplify just one of them.

- (16) *Castellazzo Bormida* (Piedmont)
- a. kwan k i maɲdʒi
 when that you.pl eat
 ‘When do you eat?’
- b. dalʼlwɔ k i drwɔmi
 where that you.pl sleep
 ‘Where do you sleep?’

Now, precisely because the complementizer is the head of an independent constituent, a left periphery can be postulated for it as well. In this schema of explanation, therefore, the *wh*-phrase that precedes *che* belongs to the left periphery of the complementizer itself, roughly as in (17). Note that the *wh*-phrase *dallwɔ* ‘where’ is simply categorized according to its relation to the predicate, as Loc(ative); we will return to the theory of left periphery elements that this labelling implies.

- (17) *Castellazzo Bormida*



The analysis in (17) predicts that the entire set of focus and topic elements should be able to appear to the left of the complementizer. Examples like those in (18)–(19) show that the left periphery of the complementizer can host not only *wh*-elements (presumably foci) but also topicalized material, including both adverbs and the lexical subject, doubled by the subject clitic following the complementizer.

- (18) *Castellazzo Bormida*
 marjo dal'lwɔ k u drwəm
 Mario where that he sleeps
 'Where does Mario sleep?'
- (19) *Cerano (Piedmont)*
 lo:/ ad'mě inda k i vĕ
 they/ tomorrow where that they go
 'Where are they going (tomorrow)?'

Our structural hypothesis therefore accounts both for data in which the complementizer has topic and focus material to its right and for data in which it has topic and focus material to its left, and thus matches the descriptive power of an articulated C field of the type in Rizzi (1997). At this point the important question arises of whether a left periphery of the type in (16)–(19) is restricted to complementizer-headed noun phrases or whether it can associate with noun phrases in general. It is evident that, to the extent that the structure of noun phrases parallels that of sentences, noun phrases are predicted to host left-peripheral material exactly as the sentence does. In fact, the left periphery of the noun phrase in (6) hosts the *wh*-item *che*, paralleling the left periphery of the sentence in (5). Similarly, the ability of the left periphery of Hungarian DPs to act as an escape hatch for a possessor phrase was the key to Szabolcsi's (1994) proposal of a parallel structure for noun phrases and sentences.

Nevertheless, an asymmetry between sentences/complementizer-headed noun phrases and other noun phrases does in fact exist – namely, that operators hosted at the left periphery of ordinary noun phrases must bind variables within the noun phrase itself. By contrast, sentences and complementizer-headed phrases can host operators binding long-distance variables. Within the analysis that we are suggesting, this asymmetry can be captured by saying that only propositions and nouns introducing propositional variables – i.e. complementizers – can support the relevant set of operators. This distinction is stipulated – i.e. it is an apparently irreducible primitive of natural languages. But exactly the same is true of Chomsky's (1973) original proposal of a C(OMP) node providing an 'escape hatch' for Subjacency in S(entences) but not in NPs, and its successors up to the present (Chomsky 1995, 2001, 2008).

1.2.1 *Combining a left periphery in the complementizer phrase and in the embedded sentence; combining two complementizers*

Suppose we accept that left-peripheral material occurring after the complementizer is associated with the embedded sentence, while left-peripheral material occurring before the complementizer is associated with the left periphery of

the complementizer itself. We then predict that the two left peripheries should combine. The simplest verification of this prediction is that topics can not only precede the cluster of *wh*-phrase and *che*, as in (18)–(19), but they can also follow it, as in (20). In present terms, this means that in (20) the *wh*-phrase occurs in the left periphery of the complementizer, while the topic appears in the left periphery of the embedded sentence.⁷

- (20) *Castellazzo Bormida*
 dal'lwə kə rə matə rə drəm
 where that the girl she sleeps
 'Where is the girl sleeping?'

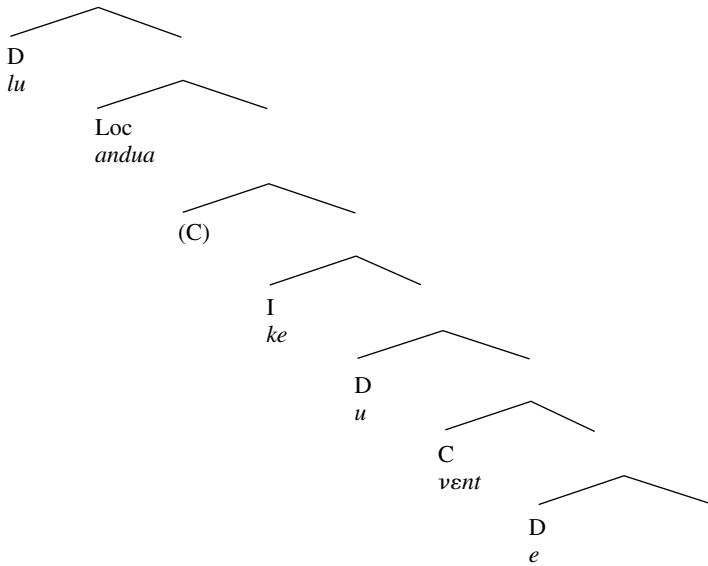
Furthermore, a left periphery associated with the complementizer, in the shape of a *wh*-phrase and/or topics preceding it, can be combined with a left periphery in the embedded sentence in the shape of a verb in C. In other words, it is possible to embed V2 (as revealed by subject–verb inversion) under a sequence of *wh*-phrase and *che* complementizer, as in (21)–(22). Note that in the *Cantoira* example (22b), the complementizer is preceded not only by the *wh*-phrase, but also by the topicalized subject.

- (21) *Mezzenile* (Piedmont)
 a. əndua k u wənt-i
 where that they go-they
 'Where are they going?'
 b. kaŋ k u vɨjunt-i
 when that they come-they
 'When are they coming?'
- (22) *Cantoira* (Piedmont)
 a. ənduə k u dyərt-e
 where that they sleep-they
 'Where do they sleep?'
 b. lu: andua k u vɛnt-e
 they where that they go-they
 'Where are they going?'

Data of the type in (21)–(22) are associated by the present theory with structures of the type in (23), in which the left periphery of the complementizer combines with the left periphery of the sentence embedded under it. The position of the verb, although we keep the conventional C label for it, is now characterized by exclusively verbal properties. Note also that two copies of the subject clitic D are present in (23), on either side of the verb; we assume that while the inverted subject clitic is in the ordinary subject (clitic) position

immediately above I, the higher copy is in the C domain.⁸ In keeping with the labelling introduced in connection with (17), the *wh*-phrase is categorized as Loc(ative) by its intrinsic content – as is the topicalized subject, notated as D like its clitic counterparts. We shall return to the lack of Top (Focus etc.) labels in section 1.3.

(23) *Cantoira*



Poletto (2000), working essentially within Rizzi's (1997) framework, argues that in sentences like (21)–(22), *che* is inserted in an intermediate position of the C field. This is distinct from the higher C position (Force for Rizzi (1997)) involved in examples like (11)–(15), and from the lower C position (finiteness for Rizzi (1997)) involved in (16)–(20). Thus, theories of CP recursion can match the descriptive power of the present theory (with a single complementizer position combined with two left peripheries) by postulating three separate complementizer positions. Despite this apparent descriptive equivalence, the two theories are clearly different.

Consider the fact that in several varieties interrogatives introduced by a *wh*-phrase alternate with interrogatives introduced by a *wh*-phrase and *che*. In at least some of them, the former have subject clitic inversion – i.e. V2,

as in (24a' and b') – but the latter don't, as in (24a and b). In many more varieties, for instance *Castellazzo Bormida*, *wh*-questions introduced by the *wh*-phrase and *che* without inversion, as in (17), alternate with yes–no questions with inversion, as in (25). Both facts lead to the conclusion that in the relevant languages the cluster of *wh*-phrase and *che* is in complementary distribution with V2, which is otherwise required by interrogatives.

(24) *Viguzzolo* (Piedmont)

- a. in'de k u drɔmæ
 where that he sleeps
 'Where does he sleep?'
- a'. ind u drɔmæ-l
 where he sleeps-he
 'Where does he sleep?'
- b. kwənt k u drɔmæ
 how.much that he sleeps
 'How much does he sleep?'
- b'. kwand u drɔmæ-l
 when he sleeps-he
 'When does he sleep?'

(25) *Castellazzo Bormida*

- rɔ drwəm-rɔ
 she sleep-she
 'Does she sleep?'

A theory such as Poletto (2000) predicts the data in (24) by assuming that the *che* complementizer sits in the lowest C position; if so, verb movement will not be able to target that position – nor, according to Poletto (2000), will it be able to target any higher position because of minimality. However, under the schema of explanation adopted here, the *wh*-phrase is in the left periphery of the complementizer *che* and the verb is in the C position of the embedded sentence, exactly as discussed for (23). Hence the two positions are different, and neither interferes with the movement paths of the other, so the fact that they cannot both be filled in (24) must be explained on other grounds.

In present terms, in varieties like (24), when the *wh*-phrase is introduced in the left periphery of the sentence, interrogative modality is lexicalized by the verb in C, as in (24a' and b'). By contrast, if the *wh*-phrase is introduced in the left periphery of the complementizer, V2 is no longer necessary and is in fact excluded, as in (24a and b); we assume that this is so because of a selectional constraint, whereby the complementizer selects the declarative

modality (i.e. the V in I rather than in C) in the embedded sentence. In the (rarer) languages of the type in (23), interrogative modality is lexicalized by the verb in C even when the sentence is embedded under the complementizer. There is no need, and indeed no evidence, for a different categorization of the complementizer which ultimately yields its different position, as in Poletto (2000).

Let us mention also that, along with languages which form interrogatives with a *wh*-phrase and *che* (no V2), as in (16)–(20), languages which form them with a *wh*-phrase and V2 (no *che*), as in (24a' and b'), and languages that form them with both *che* and V2, as in (21)–(23), there are languages that form them with neither. Thus, in (26) the *wh*-phrase introduces an interrogative sentence alone, i.e. without *che*-type complementizer, while at the same time the verb does not realize interrogative modality in C, but remains in I, as can be seen by the lack of inversion with the subject (clitic).

- (26) *Filattiera* (Tuscany)
 ke kamiza t yi
 what shirt you want
 'Which shirt do you want?'

Our examples so far involve main sentences. Leaving aside embedded questions introduced by the 'if' complementizer (to which we return in chapter 2), in embedded *wh*-questions V2 is generally not found. This is the case in languages in which no inversion is found in main *wh*-questions, as in the *Castellazzo Bormida* example in (16'), in languages that do have inversion, as in the *Cantoira* example in (22'), and also in the *Viguzzolo* example in (24') in the absence of a complementizer. In other words, V2 is generally limited to root contexts. We conclude that, contrary to the Romansh/Ladin varieties with embedded V2 in (14)–(15), matrix predicates in Northern Italian languages do not select for V in C in the embedded sentence – nor for properties of the *che*-type complementizer that in turn select for embedded V2.

- (16') *Castellazzo Bormida*
 di-m ki k ir viəŋ
 tell-me who that he comes
 'Tell me who is coming'

- (22') *Cantoira*
 di-me ki (k) e vint
 tell-me who that he comes
 'Tell me who comes'

(24') *Viguzzolo*

a ɲ sɔ k a fəŋ
 I not know what they do
 'I don't know what they do'

At the same, there is no structural impossibility of having V2 in embedded questions – a fact which is duly reflected by the existence of data like (27) attesting the possibility of the relevant patterns. In this case we assume that the verb indeed lexicalizes interrogative modality in C. It is this interpretive property that discriminates between the pattern in (16'), (22') and (24'), and the (much rarer) one in (27). From a strictly syntactic point of view, they are equally possible.

(27) *Castiglione d'Adda* (Lombardy)

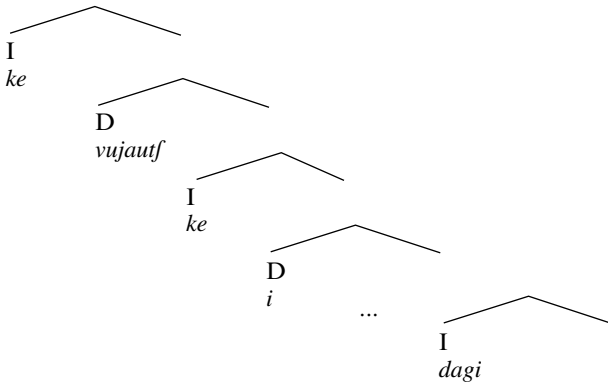
a. di-m sa tʃam-ot
 tell-me who call-you
 'Tell me who you are calling'
 b. di m sa se -t a dre a fa
 tell me what are you in the process of doing
 'Tell me what you are doing'

A final set of examples which have been considered in the literature in the context of Rizzi's (1997) proposals concerning an articulated left periphery involve the lexicalization of left-peripheral material, typically a topic, between two copies of the same *che*-type complementizer, as in (28). Examples like (28) are predicted within a theory of complementizers as functional heads, on the assumption that more than one complementizer position in the hierarchy is instantiated, along the lines of Paoli (2007).

(28) *Castellazzo Bormida*

a. l ε miɛi ke nui k a l lavu
 it is better that we that we it wash
 'It is better that we wash it'
 b. l ε miɛi ke vujautʃ k i m la dagi
 it is better that you.pl that you.pl me it give
 'It is better that you give it to me'

In the structures that we have laid out so far there is no room for two complementizers co-occurring. This is not to say that structures cannot be provided for (28). On the contrary, the relevant examples can be adequately described through recursion of the complementizer phrase. In other words, as shown in (29), the matrix predicate takes as a complement the higher complementizer, which in turn selects the lower complementizer, hosting the topic in its left periphery, and embedding the complement sentence.⁹

(29) *Castellazzo Bormida*

Needless to say, the question is not only whether a structure, say (29), can be assigned to a sentence like (28), but also what forces this structure. McCloskey (2004) entertains two alternatives. The first is that double complementizer structures semantically differ from simple complementizer structures; the other possibility is that the higher layer of complementizer structure ‘exists solely to facilitate the adjunction’ of the embedded topic material (McCloskey 2004: fn. 30). As it turns out, Paoli (2007) argues that both possibilities are instantiated in Romance. Thus, she finds that the double complementizer structures of Turinese (but ‘not the “mainstream” type spoke in Turin’) and of Ligurian are semantically restricted, in that the lower complementizer must select the subjunctive. By contrast, our data from *Castellazzo* in (28) display the double complementizer phenomenon both with the indicative in (28a) and the subjunctive in (28b). In this respect they seem entirely comparable to the data that Paoli (2007) quotes from Medieval Romance (Tuscan and other varieties).

Since the evidence we have does not point to a semantic distinction between double and simple complementizer structures, they must be distinguished on structural grounds. Specifically, based on the evidence that we have, we conclude that matrix predicates in *Castellazzo* can immediately embed complementizers, as in (29), or *wh*-phrases as in (16’), but not topic/focus material. This state of affairs can be captured in terms of a selectional constraint imposed by the matrix verb on any complementizer phrase it embeds – roughly to the effect that it can contain only clause-typing material (in the sense of Cheng (1991)), i.e. only declarative *che* or the interrogative cluster of *wh*-phrase and

che. Two alternatives are then open for the insertion of focus/topic material. First, one can position such material in the left periphery of the embedded sentence, as in (20). Alternatively, it can be positioned in the left periphery of the embedded complementizer – but then a higher complementizer must be added, satisfying the selectional constraint just defined, as in (29). In this respect, therefore, we reject the view of Paoli (2007: 1075), according to whom the lower complementizer ‘is not a subordinating particle, but overt realization of Top^o’ in medieval Romance.¹⁰

1.2.2 *Some potential problems*

Before we proceed with our discussion, we will try to clear the ground of a number of quite general questions that may be raised against our approach. Perhaps the most basic such question involves selection of an embedded complement by a matrix predicate. If a complementizer head in Romance is nominal and its projection is a noun phrase, how can we state the distinction between selecting for a complementizer phrase, i.e. a sentence, and selecting for an ordinary noun phrase? The answer is that in terms of the interpretive categories adopted here, complementizers (like bare sentences) correspond to propositions, while conventional noun phrases correspond to individual terms. Therefore selection can adequately be stated on the basis of interpretive categories.¹¹

Selection may even provide an argument in favour of the present articulation of the left periphery, as opposed to the cartographic one. Suppose that a higher predicate embeds a sentence whose leftmost and highest element is, say, a topic, as would be the case, for instance, in (18) and (19). Under Rizzi’s (1997) theory the predicate effectively embeds a topic phrase – which, according to Newmeyer (2005), provides no clear grounds for selecting the interrogative force in lower functional projections. Whether this turns out to be a serious problem or not, it does not arise in the theory that we are building. A verb selecting for a complementizer-headed noun phrase selects for the properties of the complementizer head. Alternatively, if the higher predicate selects a bare sentence – i.e. one without a complementizer – it selects for properties of the embedded verbal head.

Another property that has been consistently used to explain differences in distribution between noun phrases and sentences is Case. In the present framework, Case cannot be used to this end precisely because complementizers, which introduce sentences, are nominal – and should therefore have the same Case properties as ordinary nouns. The evidence with which Stowell (1981) introduces his Case Resistance Principle concerns the fact that noun phrases, including gerunds, can be the object of a preposition, but sentences cannot, as in (30a) vs. (30b).

- (30) a. We talked about the Marines going to Iraq
 b. *We talked about that the Marines went to Iraq

Now, elements such as *before*, *after* and *without* would seem to be prepositions since they can select noun phrases. But their Italian counterparts also embed sentences introduced by *che*, as shown in (31). Hence there appears to be no general selectional constraint against sentences as objects of prepositions or against complementizers in such a position. To the extent that more specific constraints are real, they can be stated again in terms of the interpretive notions of individual vs. propositional variable.

- (31) a. Sono arrivato prima/dopo che sei partito
 I.am arrived before/after that you.are left
 'I arrived before/after you left'
 b. Me ne sono andato senza che te ne accorgessi
 me away am gone without that you of.it noticed
 'I went without you noticing it'

A different kind of question has to do with the fact that theories in which a sentential complement is contained within a nominal layer have been proposed before – but typically differentiate between various types of sentential complements. Specifically, Kiparsky and Kiparsky (1970) differentiate complements of factive verbs, which are endowed with a nominal layer, from complements of non-factive verbs, which do not have such a layer. Crucially, different syntactic behaviours are argued to follow from the different structures. Thus, it is suggested that factive verbs are islands for extractions, since their structure makes them into complex NPs. Our judgement for Italian is that there really is no difference between extraction from a non-factive context like (32a) and extraction from a factive one like (32b) – and even extraction from 'the fact that' in (32c) is not severely degraded. In contrast, there is a strong contrast with extraction from a relative clause, as in (32d).

- (32) a. Chi pensi che non abbiamo visto?
 Who you.think that not we.have seen
 'Who do you think that we didn't see?'
 b. Chi ti dispiace che non possiamo vedere?
 who you regrets that not we.can see
 'Who do you regret that we can't see?'
 c. Chi ti dispiace il fatto che non possiamo vedere?
 who you regrets the fact that not we.can see
 'Who do you regret the fact that we can't see?'

- d. *Chi ti dispiace il motivo per cui non possiamo vedere
 who you regrets the motive for which not we.can see
 ‘Who do you regret the reason why we can’t see?’
- d’. Chi ti chiedi perchè non abbiano potuto vedere?
 who you ask why not they.have been.able to.see
 ‘Who do you wonder why they haven’t been able to see?’

The fact that (32b) patterns with (32a) may simplify our task with respect to the distinction between factive and non-factive complements, since we may not need to worry about providing different structures for them. But the clear contrast between (32a) and (32d) raises the question of whether they shouldn’t both be blocked as instances of complex NP islands. In fact, this question can be reduced to the one we discussed at the end of [section 1.2.1](#) concerning the admissibility of long-distance operator material in the left periphery of sentences vs. noun phrases. There we concluded that complementizers and sentences (because of their common propositional content) differ from other noun phrases in being able to host such material – i.e. in behaving like ‘escape hatches’. If so, we expect that ordinary noun phrases may give rise to a complex NP island effect, whereas complementizers do not. In other words, there is nothing in the present proposal contradicting standard accounts of islands in terms of subjacency/phases – though we remain strictly non-committal with respect to them.¹²

Italian does display some limited sensitivity to factive islands with adjuncts, which can have both matrix and embedded scope in sentences like (33a), while in (33b) embedded scope appears to be quite hard to obtain. This suggests that the adjunct cannot be extracted from the factive sentence – though this extraction is possible from the non-factive one. In other words, factives show an asymmetry between arguments, as in (32b) and adjuncts, as in (33b) of the type well known from Cinque (1990), and accounted for since Rizzi (1990) as a Minimality effect.

- (33) a. Perchè pensi che siano venuti?
 why you.think that they.are come
 ‘Why do you think that they are coming?’
- b. Perchè ti dispiace che siano venuti?
 why you regrets that they.are come
 ‘Why do you regret that they came?’

Now, we know that in some languages the factive vs. non-factive distinction involves the choice of different complementizers, one such language being Greek, as studied by Roussou (1994). Therefore it is natural to propose that the embedded complementizer in (33b) has some property (selected by the higher predicate) that the complementizer in (33a) does not have, for instance a

definiteness property, which, as suggested by Roussou (1994), is consistent with the interpretation of factivity. It is reasonable to think that this property triggers a Minimality effect in (33b) which is absent from (33a). In other words, the present analysis of complementizers need not interfere with current theories of extraction (with respect to which we remain non-committal, as before). Rather, the two issues are orthogonal.

Summing up so far, the structure that we propose for complementizers is motivated in section 1.1 on the basis of their nominal nature in Romance, as seen in the fact that they have the same form as *wh*-elements. Section 1.2.1 shows that this structure is compatible with the distribution of the complementizer with respect to other elements of the so-called left periphery. But note that this latter result depends on the complementizer being the head of its own projection – and does not depend on it having the same form as the *wh*-operator or even a nominal nature. In other words, though we argued that the nominal, *wh*-nature of Romance complementizers requires the structure we propose for them, nothing in the structure we propose requires nominal, let alone *wh*-properties in the complementizer.

Complementizers which do not coincide with *wh*-elements can easily be seen in Romance systems. A simple example is provided in (34), from a variety from Sardinia which distinguishes the declarative complementizer *ki*, as in (34a), from *kie* ‘who’ in (34b) and *kale* ‘which’ in (34d), as well as from *itte* ‘what’ in (34c–d). Thus, though the complementizer can be assigned to the *k*-series to which ‘who’ and ‘which’ belong, it does not coincide with any *wh*-item. For languages like (34), we of course maintain the same analysis as for Italian *che*. In fact, given the morphological relatedness of *ki* to the *k*-system of *wh*-elements, a language like *Luras* can be described as having a specialized *wh*-operator for propositional variables.

(34) *Luras* (Sardinia)

- a. m ana naðu ki enis kraza
to.me they.have told that you.come tomorrow
‘They told me that you are coming tomorrow’
- b. kie 'eniði
who comes
‘Who is coming?’
- c. itte zøn fattenɟɟɛ
what they.are doing
‘What are they doing?’
- d. itte/ kale libbru t a lleaðu
what/ which book to.you he.has brought
‘What/which book did he bring you?’

We know from Germanic languages that complementizers can belong to a non-*wh*-nominal series, namely that of demonstratives, like English *that*. We maintain the idea that the complementizer introduces a propositional variable for these languages as well. On the other hand, a *wh*-complementizer of the type generally instantiated in Romance (but also in Adyge according to Caponigro and Polinsky (2008)) is very naturally construed as a lambda abstractor; the question of how best to construe the demonstrative complementizer of English is left open here. A related open question is whether the different nature of the complementizer has reflexes in the syntax and in the interpretation of Romance vs. Germanic complementation. In chapter 2 we introduce Romance complementizer systems that have no counterpart in Germanic languages known to us, possibly indicating that these are possibilities open to *wh*-complementation systems but not to demonstrative ones.

Further afield, there are languages in which so-called complementizers are verbal, rather than nominal in nature. A case in point is the Buru language of Eastern Indonesia, as discussed by Roberts and Roussou (2003) on the basis of Klamer (2000). In Buru, *fen* can function as a main verb followed by a quotation, as in (35a), or can combine with another verb of saying introducing direct speech as in (35b) or indirect speech as in (35c); the data are from Klamer (2000).

- (35) a. Nak ana-t fene “Ng-ina nau daholo”
 3sgPoss child-Nom say 1sgVoc-mother 1sgPoss bunch-head
 ‘Her child said, “Mother, the hand (of bananas) at the top of the stalk is mine”’
- b. Da prepa fen “Sira rua kaduk”
 3sg speak say 3pl two arrive
 ‘She said “The two of them came”’
- c. Da prepa fene ringe mata haik
 3sg speak say 3sg die Prf
 ‘He said that he was already dead’

Under the present approach, in all three examples in (35) *fen* can be treated as the I head of its own projection – in this case not a noun-phrase-like projection, but a verb phrase-/sentence-like projection. In the absence of other verbal specifications, *fen* is read as a main verb, as in (35a) – which we predict to be possible precisely on the basis of the fact that it heads a verb phrase-/sentence-like constituent. On the other hand, the combination of *fen* with another verb in (35b–c) recalls so-called serial verb constructions. In Manzini and Savoia (2005) we consider serial verbs in connection with constructions in Southern Italian varieties in which aspectual/modal/motion verbs with reduced or absent inflectional properties embed sentential complements, on condition that their

temporal reference and their EPP (Extended Projection Principle) arguments coincide. We conclude that these constructions involve complex predicate formation; this implies the coincidence of temporal reference and argumental structure – and in turn these properties will typically result in a lack of inflection on one of the two verbs. The same general properties seem to hold in (35b–c), in which the so-called complementizer status of *fen* can then be described in terms of complex predicate formation – again on the basis of the usual structure in which it heads a verb phrase-/sentence-like projection.

Data of the type in (35) are worth bringing up not only for their intrinsic interest – but also because of an issue that we raised in passing in section 1.1 and then abandoned in subsequent discussion: namely, that (35) and the like are routinely described in terms of processes of historical change. Thus, according to Klamer (2000:80), ‘we can explain the synchronic distributional restrictions on *fen* if we assume that historically ... the report verb *fen(e)* has developed an alternative interpretation as a quote marker *fen* ... In contexts where *fen* is preceded by another verb which reports words, thoughts, or perceptions, it has developed a complementizer interpretation’. To quote just one similar case, for Whitman (2000: 222) in Ewe ‘the categorial feature of *bé* “say” changes from V to C’.

A closer term of comparison with the present theory is provided by Roberts and Roussou’s (2003) discussion of English *that* – an element which we have cited more than once as providing evidence for the connection of complementizers with bona fide nominal heads. In a nutshell, when it comes to *that* ‘we are dealing with one and the same lexical item which can surface as either D or C’ (Roberts and Roussou 2003: 115). This analysis is compared to Davidson’s (1997 [1968]: 828–9) idea that ‘sentences in indirect discourse, as it happens, wear their logical form on their sleeves ... They consist of an expression referring to a speaker, the two place predicate “said”, and a demonstrative referring to an utterance’. According to Roberts and Roussou (2003: 113–14), in Davidson’s analysis complementizer *that* ‘is actually the demonstrative’; by contrast, ‘it is possible to argue that *that* in terms of its position in the sentence has been grammaticalized as a C element’.

In present terms, the explanation for the range of interpretations of, say, Italian *che* or English *that* has to do with conditions internal to a single grammatical competence system; in other words, there is no necessity for invoking several competence systems in a relation of historical change to one another. Specifically, no grammaticalization is implied, understood roughly as reanalysis from lexical to functional (Roberts and Roussou 2003). In this respect the present analysis is closer to the syntactically naive one of Davidson (1997

[1968]). The gist of our proposal is that Italian *che* and its Romance counterparts are nominal quantificational heads in all cases – with a syntax typical of nominal heads, i.e. that of serving as arguments of predicates, and with a semantic content typical of quantificational heads, i.e. that of introducing variables. Similarly, we surmise that both demonstrative *that* and complementizer *that* are nominal heads and arguments of verbs; the only difference is that complementizer *that* has a propositional restrictor and demonstrative *that* a nominal one.

Before proceeding, we will consider just one more issue. Nothing that we have said so far leads us to expect that complementizers appear only in embedded contexts. In fact we have already seen a complementizer occurring in matrix sentences: namely, the *che*-type element following the *wh*-phrase in (16)–(24).¹³ The *che* complementizer also introduces yes–no matrix questions in a variety like Florence in (36), though in the same variety *wh*-questions are introduced simply by a *wh*-phrase (as in standard Italian). We surmise that the insertion of *ke* in yes–no questions corresponds to the presence of a focalization bearing on the main verb, paralleling the focalization on the *wh*-constituent in *wh*-questions, but consistent with the yes–no interpretation. If so, we can maintain that the yes–no complementizer proper (i.e. what we will call the polarity complementizer in [chapter 2](#)) is the *se* ‘if’ element that occurs in embedded yes–no questions (in Florence as in standard Italian).

- (36) *Firenze* (Tuscany)
 ke lla viene la maria
 that she comes the Mary
 ‘Is Mary coming?’

Finally, matrix instantiations of the complementizer are also possible in standard Italian, subject to a modal split, since matrix subjunctives can be introduced by the complementizer, as in (37), while matrix indicatives are not. In other Romance varieties, the *che*-type complementizer can introduce ordinary indicative sentences, as reported in the literature for Provençal (Ronjat 1937: 536 ff.) and Guascon (Rohlfs 1977: 205); an example from Sardinian is provided in (38).

- (37) Che entrino
 that they.enter
 ‘Let them enter’
- (38) *Allai* (Sardinia)
 ka dʒai dɖ a ffattu
 that already it he.has done
 ‘He has already done it’

1.3 The left periphery beyond complementizers

In the discussion of distributional evidence in [section 1.2](#), we argued that this evidence is also accounted for by our analysis when tested against the alternative view of complementizers as functional categories of the verb. The argument in favour of our hypothesis does not have to do with empirical coverage per se, but rather with the way in which this empirical coverage is achieved. In our view, there are features of the present proposal that make it more conceptually perspicuous (more explanatory) than available alternatives.

In particular, the cartographic model, which aims at an empirical coverage comparable to ours, is associated with a trend towards finer-grained phrase structure. In a sense, such an approach is simply what the empirical evidence requires – and in this sense we share it. At the same time, in the specific implementation of the cartographic programme presented, say, by Rizzi (1997, 2001, 2004), the fine articulation of syntactic structure derives from the postulation of specialized hierarchies of functional categories. The model of the left periphery that we have defined here is clearly different. Thus, where Rizzi (1997) has a hierarchy of functional projections for complementizers, we have a single complementizer position – and this position is not functional. Rather, on the evidence of its morphological form as well as its interpretation, the complementizer is the head of a noun phrase that selects the embedded sentence. In other words, a certain amount of the articulation of the so-called left periphery is achieved through the recursion of predicate–argument structures – the superordinate verb selecting the complementizer noun phrase as its argument, and this in turn selecting the embedded sentence as its argument. Critics of cartography note that functional hierarchies are potentially unrestricted devices, since a new position or set of positions in the hierarchy can always be introduced to meet new empirical evidence. The present approach exempts at least complementizer structures from this potential problem.

On the other hand, the present chapter so far only deals with the complementizer itself. Nothing that we have said touches on the independent issue of whether there is a single (conventionally C) position to which the verb can move or more than one. In previous work we have sided in favour of more than one verb position above I – based notably on the comparison of finite verbs in questions and V2 with imperatives and infinitives. Evidence concerning their distribution with respect to clitics (enclisis vs. proclisis), negation and other material leads us to the conclusion that at least two different (conventionally C) positions of the verb are involved. In adopting multiple head positions above I, we therefore follow Rizzi (1997) and the related cartographic literature.

At the same time, in the model of Rizzi (1997, 2001, 2004), the possible head positions above I are also used to order topic, focus and other phrasal material, for it is one of the postulates of standard cartography, based on Kayne's (1994) anti-symmetry, that there is a single Spec position for each head; therefore a head hierarchy automatically defines a hierarchy of phrasal material – and in fact hierarchies of phrasal material are stated as hierarchical orderings of the heads to which they attach (cf. Cinque 1999 on adverbs). The issue of the ordering of phrasal material in the left periphery is logically independent of the status of complementizers. It is evident, however, that the present reanalysis of complementizers leads us to expect that a considerable amount of currently postulated functional structure may be dispensed with in that respect as well. In this respect, potentially more promising views are held by Chomsky (1995, 2000a, 2008), according to whom each head supports any number of Specs (as for Brody (2003)), while topic, focus and similar notions correspond to interface interpretations – not to features (or categories) entering syntactic computations.

1.3.1 *Is order dictated by interpretation – or interpretation by order?*

In what follows, we concentrate on the respective position of *wh*-phrases and of lexical subjects, specifically in Northern Italian varieties with subject clitics. The topic(-like) nature of the lexical subject in Romance languages is supported precisely by their position in questions. Thus, in subject-clitic languages the subject clitic follows finite verbs in questions, including the auxiliary, as in (39). This corresponds to the position of the lexical subject in Germanic languages, and can be analysed accordingly, as a result of the clitic remaining in the same position as in declarative sentences – while the verb alternates between I in declaratives and C in questions. By contrast, lexical subjects in Romance do not appear between the auxiliary and the participle, but only after the participle or before the auxiliary, i.e. in the right or left periphery of the sentence, as in (39) again.

(39) *Oviglio* (Piedmont)

- a. ε -l am'ni marju
 is he come Mario
 'Has Mario come?'
 b. marju ε -l am'ni
 Mario is he come
 'Has Mario come?'

When *wh*-phrases are brought into the picture, it remains true that lexical subjects are generally positioned in the right or left periphery of the sentence;

however, in this latter case they are found before the *wh*-phrase, as in (40a), and not between the *wh*-phrase and the verb, as in (40b). The examples in (40c-d) illustrate the parallelism of other topics with lexical subject in this respect. Both are normally excluded from the position between the *wh*-phrase and the verb in C.

(40) *Modena* (Emilia)

- a. 'lo ki tʃam -el
 he who calls he
 'Who does he call?'
- b. *ki 'lo tʃam -el
 who he calls he
- c. la torta in'do l ε -t mesa
 the cake where it have you put
 'Where have you put the cake?'
- d. *in'do la torta l ε -t mesa
 where the cake it have you put

According to Rizzi (1997: 299), the positioning of the lexical subject is determined essentially by the *Wh*-Criterion of Rizzi (1996), whose satisfaction requires the verb and the *wh*-phrase to be in a head-Spec configuration in C(Focus)P, which forces their adjacency and hence the impossibility of a lexical subject (or other material) intervening between them. At the same time, even for Rizzi (1996: 87), the unacceptability of a sentence like (41a) in Italian contrasts with the grammaticality of (41b), in which the lexical subject intervenes between the *wh*-phrase *perchè* 'why' and the verb.

- (41) a. *Dove Gianni è andato?
 where G. is gone
 'Where has Gianni gone?'
- b. Perché Gianni è partito?
 why G. is left
 'Why has Gianni left?'

Working in a model with a single C position, Rizzi (1996) proposes that in sentences like (41b), it is the *wh*-phrase *perchè* that occupies this position, so that the verb is in the ordinary I position and the subject precedes it. However, in Northern Italian varieties the position of the lexical subject between 'why' and the verb combines with inversion of the subject clitic and the verb, as in (42), indicating that the verb is in C.

(42) *Modena*

- per'ke al pu'tein e l par'ti
 why the child is he left
 'Why has the child left?'

In fact, it is not just ‘why’ that allows for the lexical subject (or other topical material) to appear between the *wh*-phrase and the verb, as Rizzi (1996: 87) also remarks. In his terms, the obligatoriness of verb movement to C in interrogatives is weakened when a D-linked *wh*-phrase is involved, for reasons that remain unclear, so that the verb is (again) in I. Data from Northern Italian varieties confirm Rizzi’s (1996) intuition about the possibility of positioning the subject after a D-linked *wh*-phrase. At the same time, this can combine with inversion of the subject clitic after the verb, as for instance in (43), leading us to conclude that the verb is in C. Hence the phenomenon is not connected with the verb remaining in I.

(43) *Corte* (Veneto)

- a. *kel de kis marjo ljeza'ra-lo*
 which of these Mario will.read he
 ‘Which of these will Mario read?’
- b. **u'la marjo va-lo*
 where Mario goes he
 ‘Where does Mario go?’

Incidentally, note that if we combine the previous generalizations about the relative order of *wh*-phrases and lexical subjects/topics with the proposal that the complementizer is a specialized nominal head with its own left periphery, we obtain a straightforward prediction concerning interrogatives introduced by *wh*-phrase and complementizer. Namely, we predict that we will find not only the order in (23), in which the topic precedes the *wh*-phrase and complementizer sequence, but also the order in which the topic appears after the *wh*-phrase and hence before the complementizer. This prediction is confirmed by data like (44a); similar evidence is noted by Poletto and Vanelli (1995: 153) for a variety from the Turin area. Crucially, the order in (44a) appears to be restricted by the same factors that we considered for questions introduced by a simple *wh*-phrase – so that a D-linked *wh*-phrase favours the order in (44a), while a non D-linked *wh*-phrase tends to exclude it, as in (44b, b¹).

(44) *S.Maria Maggiore* (Piedmont)

- a. *kwal ad kwi gu'it ul dʒuan k u tʃama*
 which of those children the John that he calls
 ‘Which of those children does John call?’
- b. *ki k a tʃama lu:r*
 who that they call they
 ‘Who do they call?’
- b¹. **ki lu:r k a tʃama*
 who they that they call

Proceeding now with the theoretical discussion, a new take on the ‘why’ problem is offered by Rizzi (2001). In addition to the three C positions (finiteness, Focus and Force, cf. fn. 3–5) proposed by Rizzi (1997), Rizzi (2001) introduces a further C position, namely Int(errogative), located between the highest C(Force) position and the C(Focus) one. This C(Interrogative) position is meant to host elements like ‘why’ as well as interrogative ‘if’, *se* in Italian. Crucially, in Rizzi’s (1997) model, Top phrases can be freely interleaved between any C projections – therefore topicalized material is predicted to occur between ‘why’ in C(Interrogative) and the verb, even if the latter is in C(Focus). Note that an eventual unification of ‘why’ with D-linked *wh*-phrases requires the latter to be moved to (or inserted in) Spec of C(Interrogative) as well.

Abstracting away from the theoretical postulates of the cartographic approach – essentially the idea that all phrasal and head movement must be supported by the presence of a matching functional head – the empirical generalization seems to be that *wh*-phrases can move into two different positions, namely a slightly higher one that can precede a topic and a lower one that cannot. The question is whether the codification of this generalization in terms of functional structure yields predictions or insights that are otherwise unavailable. Incidentally, the labels ‘interrogative’ and ‘focus’ proposed by Rizzi (2001) for the relevant positions, while clearly interpretive in nature, are not sufficient to characterize the interpretation in full, for there is obviously no sense in which the *wh*-phrase is a Focus in one case but not in the other (or interrogative in one case but not in the other).

As already mentioned, the alternative is essentially the one suggested by Chomsky (1995, 2000a, 2008), namely that Topic, Focus and the like simply name interpretations of left-peripheral material. Suppose that the left periphery considered in (40)–(43) is freely ordered above the verb – in the absence of any functional hierarchy (of categories or features) forcing it. If so, sentences of the type in (41a), which represent the crucial case of ill-formedness, cannot be excluded on formal, computational grounds; on the contrary, their ill-formedness must be interpretive in nature.

It seems to us that a potential argument in favour of this theoretical stance comes from the fact that the unacceptability of sentences like (41a) is far from a matter of absolute judgement in particular, there is no connection between the intrinsic lexical shape of the *wh*-phrase and the range of positions that the lexical subject or other topic material can take with respect to it. Thus, Benincà (2001) notes the acceptability of sentences like (45), in which the *wh*-phrase does not appear to be D-linked. Benincà (2001) further connects the well-formedness of (45) to a ‘rhetorical question’ interpretation, implying

‘a negative answer’. We are not sure that this is the correct characterization of rhetorical questions, which seem to correspond more generally to questions that presuppose a certain answer, not necessarily a negative one (‘Who wants more money? Everybody does!’). Even so, it seems to us that (45) need not be interpreted as a rhetorical question at all – rather, it can be a genuine request for information.

- (45) Chi la fisica la sa bene?
 Who the physics it knows well
 ‘Who knows physics well?’

Independently of what its correct characterization may be, we may agree that the interpretation of (45) differs from that of (41a). This leads us to predict that in the out of the blue context implied by (41a), the utterance in (45) is equally ill-formed, while (41a) becomes grammatical if the range of interpretations that make (45) grammatical is forced. In a theory where the relative position of *wh*-phrases and topics/lexical subjects is governed by functional hierarchies of categories/features, we must assume that the very same *wh*-phrases can be associated with one or the other of these functional specifications and can be placed according to them. The other way of thinking about the relevant data is that the computational component allows for any positioning of the relevant elements (in the left periphery); however, crucially, different orderings yield different interpretations.

To the extent that both views can be used to characterize contrasts like those considered here, they are notational variants. However, one of them is arguably simpler – in more than one respect. To begin with, notions such as topic and focus are encoded only once (as interpretations) in the theory we are upholding here – while they are (redundantly) encoded twice (as interpretations and as grammatical properties) in the alternative theory. This simplicity argument has a counterpart when it comes to the single lexical entries. Assuming the maximally restrictive Inclusiveness principle of Chomsky (1995), according to which only intrinsic properties of lexical items (i.e. properties associated with that item in the lexicon) enter syntactic computation, we would have to admit that topic, focus and similar properties can be (optionally) associated with any nominal head. The alternative is to consider that these notions, like other notions that are relational in nature, do not correspond to features/categories at all, but rather to configurations (cf. Chomsky 2000a on theta-roles) – defined in this case at the LF interface. These simplicity considerations, in the absence of empirical evidence, prove decisive in our view.¹⁴

1.3.2 *Embedded contexts*

In the preceding section, we considered the relative order of *wh*-phrases and lexical subjects in matrix interrogatives. A further element of complexity is introduced by embedded interrogatives. Here the judgement of Rizzi (1997: 289) is that in Italian the topic can precede the *wh*-phrase, as in (46a), while the configuration in (46b), in which the topic follows the *wh*-phrase, is 'slightly marginal'. For us, both examples are equally well-formed. In fact, Rizzi's (1997) structural schemas allow the topic to either precede or follow the *wh*-phrase, assuming its position to be the same in (46a) and in (46b), namely C(Focus).

- (46) a. Mi domando il premio Nobel a chi lo potrebbero dare
 myself I.ask the prize Nobel to whom it they.could give
 'I wonder to whom they could give the Nobel prize'
- b. Mi domando a chi il premio Nobel lo potrebbero dare
 myself I.ask to whom the prize Nobel it they.could give

Recall from the previous section that for Rizzi (1997), movement of the verb to C(Focus) normally excludes the *wh*-phrase–topic order in matrix questions. Therefore he automatically predicts that the latter resurfaces in embedded questions, as in (46b), because of the absence of verb movement. Data from Northern Italian varieties, however, call his explanation into question. In a null-subject language like Italian, it is hard to detect the position of the verb in the sentence on independent grounds, but in Northern Italian varieties the position of the subject clitic provides a reliable independent test. In many of these languages, inversion of the verb with the subject clitic does not take place in either matrix or embedded questions, providing evidence that the verb maintains its I position in all cases. Yet in the same languages, lexical subjects and other topics may appear in either right-peripheral or left-peripheral positions in matrix interrogatives, while the occurrence of the topic or lexical subject between the *wh*-phrase and the verb is not attested. This contrasts with embedded questions, in which the lexical subject/topic can either precede or follow the *wh*-phrase, as shown in (47).

- (47) *Fontanigorda* (Liguria)
- a. kølu li duve u druome
 that.one there where he sleeps
 'Where does he sleep?'
- b. ne suo kølu li duve u druome
 not I.know that.one there where he sleeps
 'I don't know where he sleeps'

- b'. ne suo duve kølu li u druome
 not I.know where that.one there he sleeps
 'I don't know where he sleeps'

In other words, the contrast between matrix and embedded questions in a language without subject-clitic inversion is exactly the same as the one observed in a language with subject-clitic inversion, as in (48) vs. (40).

(48) *Modena*

- a. a n sɔ briza to fra'del in'do al va
 I not know not your brother where he goes
 'I don't know where your brother goes'
- b. a n sɔ briza ki lo l a tʃa'mɛ
 I not know not who he him has called
 'I don't know who he has called'

The conclusion that the ordering of the lexical subject/topic with respect to the *wh*-phrase in matrix questions is not determined by the position of the verb with respect to the *wh*-phrase is supported by at least one independent argument, from subject-clitic languages with interrogative inversion. In several relevant Northern Italian varieties, the subject clitic is not simply inverted after the verb but doubled before and after the verb. The relevant examples are (23) for questions introduced by a combination of *wh*-phrase and complementizer and (24a', b') for questions introduced by a simple *wh*-phrase. Therefore the supposed Spec-head adjacency of the *wh*-phrase and the verb is systematically disrupted by the preverbal subject clitic. In terms of the structure already provided in (23), in (24a', b') a D position intervenes between the *wh*-phrase in the left periphery of the sentence and the verb in C. If one wanted to claim that the clitic represented adjoined material, one would at the very least have to clarify how the double adjunction (of a proclitic and an enclitic) comes about.

In short, for the various reasons reviewed, the contrasts relating to the positioning of the lexical subject in matrix and embedded questions cannot be due to the *Wh*-Criterion (or its variants). In other words, the relative ordering of *wh*-phrases and topics/lexical subjects seems to be determined by their intrinsic properties, without the position of the verb playing any role. If so, then it is no longer clear that the general acceptability of *wh*-phrases in front of lexical subjects/topics in embedded questions, and their acceptability in certain matrix questions (introduced by 'why', D-linked *wh*-phrases etc.) should not be accounted for in the same way.

Answering the question of what would unify these various contexts is beyond the scope of the present chapter, which aims simply at establishing a

theory of complementizers – including of course its compatibility with a more general model of the left periphery. Nevertheless, the solution envisaged by Manzini and Savoia (2005) is worth mentioning, since it represents an extension of sorts of the approach to complementizers themselves. The idea is that embedded interrogatives, in which the *wh*-phrase can precede topical material, are effectively (free) relatives, of which the *wh*-phrase represents the nominal head. In other words, just as the complementizer is not in the left periphery of the sentence but rather projects its own noun phrase, the *wh*-phrase can project its own noun phrase, which embeds a sentence, eventually including topical material.

This proposal can equally be applied to matrix questions, as Manzini and Savoia (2005) effectively do for matrix questions introduced by ‘why’. In other words, ‘why’ is not in the left periphery of interrogative questions; rather, it systematically introduces them as a sort of propositional operator (complementizer). In any event, we may assume that ‘why’ projects a nominal constituent – which could then be taken to be more akin to a (free) relative. The latter proposal could further be extended to D-linked *wh*-phrases, and more generally to all *wh*-phrases that embed a topic.

Other extant proposals in the literature seem to rest on intuitions compatible with the analysis just sketched. We have already mentioned Rizzi’s (1997) treatment of ‘why’ as a complementizer head. Kayne and Pollock’s (2001) discussion of *pourquoi* ‘why’ in French is also relevant. They consider the fact that ‘why’ and other *wh*-phrases like *en quel sens* ‘in which sense’ do not trigger inversion of the verb with the subject clitic in French matrix questions, in contrast to other *wh*-phrases. What they suggest is that questions like *en quel sens les fleurs parlent* ‘in which sense flowers speak?’ ‘include an abstract verb corresponding to *say*’, as if one were to say ‘In which sense are you saying that flowers speak?’. This proposal shares with ours the intuition that ‘why’ and other items of the same class belong to a different sentential(-like) constituent with respect to the sentence they introduce.

At the same time, for reasons of restrictiveness of the theory, but also for strictly empirical reasons, we reject what Kayne (2006, 2008a) calls silent categories (Manzini and Savoia 2008a, 2009a, 2010; Savoia and Manzini 2010), including the ‘abstract verb’ of Kayne and Pollock (2001). In our intuition, the right interpretation for a sentence like (45), with the order *wh*-phrase–topic, is more likely to be simply ‘who (is it that) knows physics?’. Similarly, the interpretation of *en quel sens les fleurs parlent?* could be reconstructed as ‘in which sense (is it that) flowers speak?’. In other words, in both cases the embedded sentence is predicated of the *wh*-phrase – more or less as we expect for (free) relatives, or clefts.¹⁵

For the sake of completeness, we note that in the discussion that precedes, the position of lexical subjects is assimilated to that of topics. However, this does not necessarily hold in all languages (witness the case of English), nor even in all Romance languages. Thus, in French, the lexical subject appears between the *wh*-phrase and the verb in so-called complex inversion examples, as in (49), independently of interpretive factors. The fact that the verb is inverted with respect to the subject clitic means that it is in a relatively high (conventionally C) position, so that the lexical subject must be in a higher position still – yet not a topic position. The present analysis already provides a non-topic position for subjects above C, namely the position taken by the preverbal subject clitic in the structure in (23). The ability of the lexical subject in French to appear in a position which in Northern Italian varieties is only available for subject clitics correlates with the fact that while lexical subjects are obligatorily doubled by subject clitics in Northern Italian varieties, they lexicalize the subject alone in French – and determine complementary distribution with subject clitics.

- (49) Quand Jean vient-il?
 When J. comes-he
 'When is John coming?'

At the same time, French, like Northern Italian varieties, reserves the lowest subject position, internal to the I domain, for clitics; therefore only subject clitics can be found inverted after the verb in questions. This contrasts with a language like English, in which the lexical subject ordinarily appears after the verb in questions – and hence presumably occupies the same position as subject clitics in Romance languages. The English-type position of the subject can also be seen in some Romance languages, namely Romansh varieties in which the lexical subject can appear between the auxiliary and the participle, as in (50).¹⁶

- (50) *Trun* (Grisons)
 a. ain ilts ta'ħɔ:R kurdai pɛR tiara
 are the dishes fallen to ground
 'Have the dishes fallen to the ground?'
 b. an iħts a'fɔnts dur'miu
 have the children slept
 'Have the children slept?'

There is a final set of data from the work of Rizzi (1997) that goes potentially unpredicted by the present approach. According to the data presented there, a topic can precede a *wh*-phrase not only in matrix questions, but also in embedded questions. However, a *che*-type complementizer cannot be preceded by a topic. The relevant data for embedded questions are of the type in (46);

by contrast, Rizzi (1997) judges (51) and the like to be ill-formed. Benincà (2001), however, does not exclude topics in front of declarative complementizers altogether, since according to her an example like (51b) is well-formed. However, (51b) involves a so-called ‘hanging topic’; for Benincà (2001), so-called clitic left dislocation, as exemplified in (51c), remains ill-formed. For us, the two sentences in (51b–c) have the same status; i.e. they are both acceptable. The same holds for a sentence like (51a), which is ambiguous between the descriptive categories of hanging topic and clitic left dislocation.

- (51) a. Credo il tuo libro che loro lo apprezzerebbero molto
 I-believe your book that they it would.appreciate a.lot
 ‘I believe that they would appreciate your book a lot’
- b. Sono certa questo libro che non ne ha mai parlato nessuno
 I.am certain this book that not of.it has ever spoken anybody
 ‘I am certain that nobody has ever talked about this book’
- c. Sono certa su questo tavolo che non ci hanno messo niente
 I.am certain on this table that not there they.have put anything
 ‘I am certain that nobody put anything on this table’

The theory of Rizzi (1997) is constructed in such a way as to exclude examples of the type in (51). In particular, the *che* complementizer in these examples is identified with the highest C position, namely C(Force), which closes off the C field (cf. fn. 4, 6); therefore the prediction is that no left-peripheral material can precede it. In the present approach, however, in which *che* is the head of an independent nominal projection, the possibility must be open for its left periphery to host topical material. In fact, this is the structure we have postulated for sentences like (23) in section 1.2. Therefore examples like (51) are predicted to be grammatical.

1.4 Conclusions

In a nutshell, the present proposal represents a viable alternative to theories of C as a functional projection of the verb – and possibly a better one in that it allows us to simplify functional architectures, treating complementizers in terms of the recursion of predicate argument structures. Another advantage resides in the simplification of the lexicon, to the extent that the complementizer can be given a unified lexical entry with the *wh*-phrases with which it is (often) homophonous. Though this would appear to be an even smaller gain than the previous one, consider that in a minimalist model the lexicon is all there is to language variation – therefore simplifications in the lexicon are simplifications of the only learning task children have in front of them.

In the next chapter, we will pursue this second theme, providing some case studies in language variation and illustrating the account available for them under the present model. We argue that the model we are proposing is capable of accounting for the intricate parametrization of Romance varieties in a transparent way, effectively predicting that certain parametric values should be instantiated. In our view, these predictions depend on treating the complementizer as an argument, and specifically as an argument belonging to the same *wh*-series as classical *wh*-phrases. The burden of proof is on other theories to show that the same results can be mimicked if complementizers are functional projections of the verb. This may very well be possible – but almost certainly through additional assumptions.

2 *Variation in Romance* *k-complementizer systems*

As observed in [chapter 1](#), finite complementizers in Germanic and Romance are clearly nominal, belonging to argumental series such as demonstratives and *wh*-elements respectively. Therefore we propose that the complementizer is not a functional category of the verb, but rather an independent nominal head, which satisfies an argument slot of the matrix verb and which takes the embedded sentence as its complement. The aim of the present chapter is to investigate the range of variation in the lexicalization of complementizers in Romance varieties. We argue that only the treatment of the complementizer sketched in [chapter 1](#) allows for a transparent account of the observed variation. First, the overlapping of Romance complementizers with the *wh*-system follows precise patterns, which excludes the possibility that we are merely dealing with homophony. Second, an account of the variation internal to the complementizer system requires the complementizer to be interpreted as introducing a propositional variable – which is natural if it is a nominal head, but not if it is a functional projection of the verb. In [section 2.3](#), we also consider the finiteness restriction that complementizers are subject to, arguing that they do not select for embedded temporal/modal properties (this being a potential argument in favour of their status as a functional projection of the verb), rather they select for the properties of the embedded EPP argument – hence of the proposition as a whole.

2.1 **Systems with two *k*-complementizers**

The system of standard Italian, which we have considered in [chapter 1](#), is characterized by a single complementizer introducing finite declarative clauses – which is also the distribution familiar for English *that*. However, there are many Central and Southern Italian varieties which have two finite declarative complementizers, as in (1)–(2). One of the two complementizers, generally *ka*, is systematically found to introduce complements to verbs

of ‘saying’ and ‘knowing’. The other complementizer, generally *ke/ki* (*kə*), is found to introduce complements to verbs of ‘believing’ and ‘wanting’, though in some varieties (*Guglionesi*) some predicates alternate between the two complementizers. In other varieties, where all these predicates select the same complementizer *ka*, a second complementizer is found to introduce complements to ‘before’ and ‘after’ (*Montenerodomo*). Another important common property of these systems is that they involve some overlap between the complementizer system just described and the *wh*-quantifier system. Crucially, this overlap never affects the *ka* complementizer; instead it is the *ke/ki* complementizer that quite generally overlaps with the *wh*-quantifier for ‘what’.

(1) *Guglionesi* (Molise)

- a. i. m ɔnnə dəttə ka vɛ krɛ
 to.me they.have said that he.comes tomorrow
 ‘They told me that he will come tomorrow’
- a. ii. pɛndzə ka vɛvɛ krɛ
 I.think that he.comes tomorrow
 ‘I think that he will come tomorrow’
- b. i. vujjə kə vi krɛ
 I.want that you.come tomorrow
 ‘I want you to come tomorrow’
- b. ii. pɛndzə kə vɛvɛ krɛ
 I think that he.comes tomorrow
 ‘I think that he will come tomorrow’
- b. iii. sɔŋg aʃʃɛutə prəmə kə tteu mənəssə
 I.am gone.out before that you would.come
 ‘I went before you came’
- b. iv. sɔŋg aʃʃɛutə dɔppə kə ssi mənɛutə
 I.am gone.out after that you.are come
 ‘I went after you came’
- c. kə ffi
 what you.do
 ‘What are you doing?’

(2) *Montenerodomo* (Abruzzi)

- a. i. m omə dittə ka vi dumɔ:nə
 to.me they.have said that you.come tomorrow
 ‘They told me that you will come tomorrow’
- a. ii. pɛndzə ka issə vɛ dumɔ:nə
 I.think that he comes tomorrow
 ‘I think that he will come tomorrow’
- a. iii. vulessə ka mənɪʃʃə
 I.would.want that he.would.come
 ‘I would want him to come’

- b. i. so ʃʃeutə preimə kə mənɛivə
I.am gone.out before that he.came
'I went before he came'
- b. ii. so ʃʃeutə dopə kə tu avi mənɛutə
I.am gone.out after that you had come
'I went after you came'
- c. kə ffi
what you.do
'What are you doing?'

The existence of double complementizer systems in varieties of Central and Southern Italy has been noted in the literature. However, Rohlfs (1969 [1954]: 190) puts it in the same bracket as another form of split lexicalization for sentential introducers, found in so-called Balkan languages as well as in some Southern Italian and Sicilian varieties. The relevant systems are characterized either by the lack of morphological infinitives or by their very reduced presence. Therefore the embedded infinitival clauses of English or Italian, associated with control and raising interpretations, are rendered in these Balkan (-like) languages by the embedding of finite clauses introduced by a specialized particle, which in Calabrian varieties, for instance, is *mu*. Thus, for Rohlfs, Neapolitan *pənsə ca vənə* 'I think that he come' vs. *vògliə kə mmangə* 'I want that he eats' and Calabrian *pensu ca vèni* 'I think that he comes' vs. *vogghiu mu (mi) mangia* 'I want that he eats' represent strictly comparable systems.

The strongest argument for the independence of the two phenomena is that they actually combine. Thus, the Calabrian variety from *Arena* in (3), which has control and raising complements introduced by *mu*, also independently presents two complementizers of the *k*-series, roughly with the distribution described above for *Montenerodomo*, as in (2a–b). Of the two complementizers, it is the *ki* one, selected notably by 'before' and 'after', that has the same form as the *wh*-quantifier for 'what', as in (3c).

(3) *Arena* (Calabria)

- a. i. mi 'ðissiru ka vɛni dɔmani
to.me they.said that he.comes tomorrow
'They told me that he will come tomorrow'
- a. ii. kriju ka vɛni
I.believe that he.comes
'I think that he will come'
- a. iii. ɛ mmieʝju ka viɛni
it.is better that you.come
'It is better for you to come'
- b. neʃʃivi duɛppu ki vinni
I.went.out after that he.came
'I went after he came'

- c. ki ffatʃi ijju
 what does he
 ‘What is he doing?’
- d. vuɛjju mu niʒʃʃu
 I.want Prt I.go.out
 ‘I want to go out’

We assume that introducers of the type of *mu* in Calabrian (or *të* in Albanian) are not complementizers in the sense defined in the preceding paragraph – i.e. they are not (nominal) heads taking the embedded proposition as their (sentential) complement. Rather, they are internal to the embedded clause, to which they contribute modal properties (Rivero 1994; Roberts and Roussou 2003) and/or EPP properties involved in control and raising interpretations. Thus, we will put *mu* and the like aside without further discussion. Ledgeway (2003a, 2005, 2009) also argues against the identification of double complementizer systems with systems including a complementizer and a so-called subjunctive particle. It is possible that the three-complementizer split (*ca*, *cu* and *che*) described by Ledgeway (2005: 367 ff.) on the basis of Sgrilli (1983) for Early Salentino represents a system comparable to *Arena* in (3) – with *cu* representing the ‘subjunctive particle’, very much like *ku* in contemporary Salentino varieties.

Let us go back, then, to (1). The distribution of *ka* and *kə* in varieties such as *Guglionesi* is reminiscent of the fact that standard Italian employs the indicative in complements of ‘to say’ and the subjunctive in complements of ‘to think’ or ‘to want’, with regional variants allowing for the indicative at least under ‘to think’, as in (4).

- (4) a. Dico che viene
 I.say that he.comes
 ‘I say that he will come’
- b. Penso che venga/ viene
 I.think that he.come/ he.comes
 ‘I think that he will come’
- c. Voglio che venga
 I.want that he.come
 ‘I want him to come’

This similar distribution (and the somewhat similar distribution of indicatives vs. infinitives in the English translations) seems to suggest that some common property lies at the core of the complementizer split and the modality split. However, it should be stressed that it is not the case that one phenomenon depends on the other. In particular, in the examples in (1)–(2) the two

complementizers can embed the same indicative verbal forms, while *ka* in the *Montenerodomo* example in (2a) can embed the one form of subjunctive that appears to be present in the language, which can be roughly characterized as the counterfactual. Further examples of this pattern in *Guglionesi* are provided below in (9). In other words, there is no obligatory selection of the verb modality by the complementizer or vice versa.

The evidence at our disposal also allows us to exclude the possibility that the complementizer split can be linked to the relatively impoverished modal system of varieties like (2); in other words, not only is it the case that there is no selection relation between complementizers and subjunctives, but there is no complementary distribution (functional equivalence) either. A case in point is represented by Sardinian varieties such as *Paulilatino* in (5), in which a full-blown indicative – subjunctive system of the standard Italian type combines with a double complementizer system of the type exemplified in (2) with Abruzzese varieties.

(5) *Paulilatino* (Sardinia)

- a. i. m anta nau ka bbeni kkraza
to.me they.have said that he.comes tomorrow
'They told me that he will come tomorrow'
- a. ii. pentsɔ ka 'bbenizi
I.think that you.come
'I think that you will come'
- b. i. pentsɔ ki 'bbendzεðε
I.think that he.come
'I think he will come'
- b. ii. kerdzɔ ki 'bbendzεzε
I.want that you.come
'I want you to come'
- b. iii. sεɔ bis'siu appustis ki ze βenniu
I.am gone.out after that you.are come
'I went after you came'
- b. iv. sεɔ bis'sia primma ki esse βenniu ðue
I.am gone.out before that be come you
'I went before you came'
- c. itte fa'eze
what you.do
'What are you doing?'

It will be noted that the *ka* complementizer combines with morphological indicatives in (a) while the *ki* complementizer combines with morphological subjunctives. However, the correlation between complementizer and verb modality breaks down in connection with complements to 'before' and 'after'

in (4biii) and (4biv) respectively. Following the pattern of standard Italian, the complement of ‘before’ is a subjunctive (a counterfactual) while the complement of ‘after’ is an indicative. Nevertheless, in both cases the complement is introduced by the ‘modal’ complementizer *ki*. This should of course be taken together with the existence of languages, like *Montenerodomo* in (2) or *Arena* in (3), in which complements to ‘before’ and ‘after’ maintain a specialized complementizer while other declarative sentential complements are uniformly introduced by *ka*.

The lack of a one-to-one correlation between complementizers and modality is discussed at length by Ledgeway (2003a, 2005), who studies double complementizer systems in Early Neapolitan and other Southern Italian texts. He concludes that indicative clauses are frequently introduced by ‘CHE’. He also notices that selection for complementizers cannot be directly imputed to matrix predicates, since ‘it is not difficult to find minimal pairs ... where the same main clause predicate selects in one case for an indicative clause headed by CA and in another for an indicative clause headed by CHE’ (Ledgeway 2005: 348). In other words, the historical varieties studied by Ledgeway have a complementizer system strictly comparable to that illustrated here by *Guglionesi* in (1) and in fact also by *Paulilâtino* in (5); the similarity between Ledgeway’s varieties and Sardinian ones is independently noted by Damonte (2006), who extends Ledgeway’s analysis to them.

Nevertheless, Ledgeway subscribes to Formentin’s (1998: 432) conclusion that ‘the conjunction *ca* never introduces subjunctive clauses’, thus establishing a one-way implication between subjunctive and CHE complementizers. This one-way implication is not upheld by our data. Recall that we have already reviewed systems, such as *Montenerodomo* in (2), in which *ka* introduces a subjunctive (2aiii). More to the point, Sardinian varieties, which have fully productive subjunctives, show that both complementizers combine with both moods (indicative and subjunctive). Relevant data are found below in (11) for the variety from *Làconi*. Similarly the free alternation of *ka* and *kə* with subjunctives is exemplified for *Guglionesi* in (9). Evidence that the subjunctive does not imply *chi* independently emerges in the Baunei corpus of Damonte (2006: 92), who quotes examples of *ca* with the subjunctive such as *Mi pare ca custas cadirasa siente meda comodasa* ‘it seems to me that these chairs are (lit: be) very comfortable’.

2.1.1 *Definite and indefinite complementizers – and alternative analyses*

According to the conclusions of [chapter 1](#), the finite *k*-type complementizer in Romance languages is an independent nominal head introducing a propositional

variable, whose content is restricted by the embedded sentence. If so, the difference between the two complementizers in (1) is most naturally construed as a difference between two types of variable/quantification. We could assume, for instance, that the *ka* complementizer is a definiteness element, effectively the counterpart to a definite determiner, so that the LF of the *Guglionesi* example in (1a.ii) would be of the type in (6a). By contrast, the LF of a sentence like the *Guglionesi* example in (1b.ii) could include an indefinite quantification, corresponding to a free variable bound by existential closure, as sketched in (6b).

- (6) (*Guglionesi*)
 a. I think the *x*: *x* he comes tomorrow
 b. I think for some *x*: *x* he comes tomorrow

Because the evidence and the results discussed in this paper are mostly morphosyntactic in nature, we must content ourselves with a sketchy account of the interpretation. The evidence presented above is, however, sufficient to yield the generalization that of the two complementizers present in languages like *Guglionesi* in (1), it is always the indefinite complementizer that overlaps with a *wh*-quantifier. This is because the other complementizer effectively introduces a definite description, which is hardly compatible with *wh*-quantification – while the indefinite complementizer introduces a propositional variable subject to existential closure, more or less like the argumental *wh*-variable. Manzini and Savoia (2005) provide examples from a considerable number of languages which behave like *Guglionesi*, and in all cases the same generalization holds, as shown by the summary table reproduced in (A) in the Appendix. Of course it is also logically possible to have double complementizer systems of the type in (6) in which no overlap with the *wh*-system is found, as summarized in table (A') in the Appendix; the *Paulilätino* sentence in (5) exemplifies this type of system.¹

The preceding results are important for the present discussion for more than one reason. A preliminary point is that the identity of form shown by complementizers and *wh*-elements covers grammars that are not just different in general terms, but are specifically different in terms of the finite complementizer system – having either a single complementizer or a split between complementizers. In other words, it cannot be objected that the formal identity of the complementizer with the *wh*-system is in a sense the property of just one grammar, and therefore to be treated as accidental. More importantly, the fact that in two complementizer systems it is the indefinite complementizer that coincides with the *wh*-system supports the semantics for complementizers sketched here, and indirectly the syntax in chapter 1.

Now the question is: under the alternative view, which has complementizers as C functional heads, could one provide characterizations for the two different complementizers, and capture the overlap with *wh*-quantifiers (or lack thereof)? Ledgeway (2003a, 2005), dealing with data closely comparable with ours, does provide an answer to the first part of the question. Before evaluating it, let us note that the two complementizer systems of languages like those in (1) may have a reflex even in a language like Italian in complementizer deletion phenomena; we will therefore consider first what the existing literature says about those.

In fact there appears to be a good match between matrix predicates that require the indefinite complementizer in (6b) and those which, according to the literature, allow for the deletion of the *che* complementizer in Italian, as in (7b–d). By contrast, contexts which require the definite complementizer, as in (6a), seem to coincide with those that do not admit complementizer deletion in Italian, as illustrated in (7a).

- (7) a. So *(che) viene
 I.know that he.comes
 'I know that he is coming'
- b. Penso (che) venga
 I.think that he.come
 'I think that is coming'
- c. Vorrei (che) venisse
 I.would.want that he.came
 'I would like him to come'
- d. E' meglio (che) venga
 It.is better that he.come
 'It is better for him to come'

The issue is made more complex by the existence of at least one variety in which effectively any matrix predicate admits complementizer deletion. The latter is identified by Cocchi and Poletto (2002) with the variety spoken in Florence. As in many cases involving Italian, the question can legitimately be asked whether, on the contrary, the judgements attributed to the standard are not unduly constrained by normative considerations; in fact even the judgements in Cocchi and Poletto (2002) appear to be unnecessarily restrictive for the Italian spoken in Florence.² Be that as it may, the reason to address the (potential) pattern in (7) is that phenomena of complementizer dropping have been prominent in the theoretical literature since the study of the *that*-t filter in English by Chomsky and Lasnik (1977). Therefore distributions of the type in (7) have been discussed more than once (Poletto 2001; Giorgi and Pianesi 2004), providing us with a possible alternative to the analysis advocated here.

On the basis of the preceding discussion of double complementizer systems, we would be led to characterize the alternation in (7) by saying that in order to introduce a propositional definite description the complementizer is necessary, as in (7a); while other types of propositional complementation admit the bare embedding of a sentence under the matrix verb, as in (7b–d). As for the question of why contexts obligatorily introduced by the complementizer feature the indicative, while subjunctive contexts admit complementizer dropping, we can assume that the indicative vs. subjunctive split is itself connected with the definite or indefinite binding of a propositional variable. In fact, Manzini (2000) notices that mood selection is not exclusively determined by the matrix predicate. Thus, ‘to know’ in (8a) normally requires the indicative. The latter remains possible when the matrix predicate is questioned, as in (8b), in which case the reading is still a factive one presupposing the truth of the embedded proposition. Questioning the matrix sentence also renders the subjunctive possible, in which case, however, the embedded proposition is in the scope of the question operator, yielding a non-presuppositional reading.

- (8) a. So che è/ *sia venuto
 I.know that he.is/ he.be come
 ‘I know he came’
 b. Sai che è venuto?
 you.know that he.is come
 ‘Do you know that he came?’
 c. Sai che sia venuto?
 you.know that he.is come
 ‘Do you know if he came?’

Alternations like (8b–c) can be described by saying that the indicative (like *ka*-type complementizers) introduces a propositional definite description, read outside the scope of polarity operators like the question operator. By contrast, the subjunctive (like *ke*-type complementizers) introduces an indefinite propositional variable, interpreted within the scope of polarity operators. This accounts for the parallelism between the complementizer split and the indicative vs. subjunctive split apparently present in (7).

Let us now consider Giorgi and Pianesi (2004) and Poletto (2001). They assume, as we do, that so-called complementizer deletion does not correspond to the actual deletion of lexical material nor to the zero instantiation of the C position. For Giorgi and Pianesi (2004), while the sequence of *che* and a subjunctive provides separate instantiations for modality (through the *che* complementizer) and for agreement (through the I position of the verb), in complementizer deletion contexts, the Mood and Agr properties are conflated in the

I head, and simply lexicalized by the verb. The reason why *che* cannot delete in indicative contexts has to do with the fact that in indicative contexts T and Agr properties cannot conflate, but must be independently lexicalized. To be more precise, ‘the T-features are duplicated in *che/C* in such a way that the indexical component can be evaluated outside the clause’ (Giorgi and Pianesi 2004: 200). According to Giorgi and Pianesi (2004) the satisfaction of this latter condition is independently required in order to obtain the so-called Double Access Reading, under which the situation denoted by the embedded sentence is taken to hold both at the time of the matrix event and at the time of utterance.

What is relevant here is that Giorgi and Pianesi’s (2004) framework could in principle account for double complementizer systems, as the result of a split between exponents of C endowed with Tense (*ka*) and exponents of C endowed with Mood (*ke/ki*). But since there is no (direct) connection between the *wh* system and subjunctive Mood (e.g. *wh*-questions do not necessarily select the subjunctive) the overlapping of modal complementizers and the *wh*-system (to the exclusion of the non-modal complementizers) would not be predicted. The present discussion provides the missing link; but crucially it does so by extricating the complementizer from the Mood and Tense system to which Giorgi and Pianesi (2004) confine it – and by assuming that it is an argumental element. Correspondingly it is removed from the C position and construed as an independent (nominal) head.

Let us now turn to Poletto (2001), who draws a parallel between complementizer deletion and Germanic V-to-C movement; the latter, at least in some languages, is in complementary distribution with the lexicalization of the complementizer (and thus seems to result in its deletion). The idea is that ‘the class of verbs selecting C[omplementizer] D[ele]tion complements is exactly the same as in Germanic embedded V2 contexts’ (Poletto 2001: 267). Thus both phenomena can be captured by the movement of the verb to the C position. On the assumption that ‘CD is possible only when the embedded verb is a subjunctive, a future or a conditional form’ and that ‘these forms all have a modal quality’, Poletto (2001: 278) proposes that ‘a [–realis] feature ... is realized on the head of the complement and attracts the verb into the CP domain’.

Let us leave aside the question of whether the verb does or does not move to the C position in Italian complementizer deletion contexts, what is relevant here is Poletto’s (2001) characterization of these contexts in terms of a [–realis] feature in C. In these terms, it is not difficult to recognize the common conceptual core of Poletto’s (2001) analysis and Giorgi and Pianesi’s (2004), namely that complementizer deletion involves modal properties, absent from indicative contexts. The point on which we differ from both analyses is once again made clearer if we try to extend Poletto’s (2001) account to the double

complementizer systems considered here. One could assume, for instance, that what we have called the indefinite complementizer lexicalizes the *-realis* feature, while the other complementizer satisfies the *+realis* feature in C. But if so, there is no reason why the former should overlap with the *wh*-system (to the exclusion of the latter), since *wh*-questions do not appear to have any (immediate) connection with *-realis* (not being restricted, say, to the subjunctive). As before, the present theory has the advantage that it can make direct predictions on this point – but these crucially depend on complementizers being extricated from the C position, and treated as argumental, nominal heads.

We consider next the proposals put forth by Ledgeway (2003a, 2005). The empirical generalization that Ledgeway proposes concerning the distribution of the two complementizers depends on a further set of data concerning the distribution of CHE and CA with respect to embedded left-peripheral material. In his corpus of eleven early texts, he finds that

out of a total of 327 examples of CA-clauses ... a mere 10.1% were found to contain one or more elements in the left periphery, whilst from a total of 1,061 examples of indicative clauses introduced by CHE ... 41.8% were found to host one or more elements in the left periphery ... Secondly even if there do occur some examples of the complementizer CA preceding one or more elements of the left periphery ... with very few exceptions, all such examples involve elements of the Focus field, namely ... foci ... and fronted indefinite quantifiers. (2005: 360)

Ledgeway's characterization of CA and CHE is in fact partially based on mood selection, since he assumes that 'CA and CHE are invariably merged in ... *Fin*^o ... as an overt reflex of the different modal specifications (indicative vs. subjunctive) they check in *Fin*^o, from which automatically follows the claim that CA only occurs in indicative clauses and CHE only in subjunctive clauses' (2005: 372). However, 'the appearance of topics and foci forces the complementizers CA and CHE to move out of *Fin*^o to target *Force*^o' (2005: 376); crucially 'this movement operation is spelt out morphologically only in the case of the indicative complementizer CA which invariably surfaces as CHE' (2005: 374–5).

The empirical generalization, namely that the presence of an embedded left periphery forces the complementizer to be CHE, deserves to be investigated for contemporary varieties as well. What we have found is that our varieties do not observe such a restriction, since both foci (e.g. (9c)) and topics (including clitic left dislocation in (9b)) appear under *ka* in *Guglionesi* in (9) and in the Sardinian varieties in (10)–(11).³ We also exemplify topics and foci under *kə* and *ki* respectively, in order to establish that our varieties do not retain the asymmetry between the two complementizers, simply reversing their respective positions (i.e. *ka* higher and *kə/ki* lower).⁴

(9) *Guglionesi*

- a. m annə ðəttə ka krɛ mənəvə
to.me they.have said that tomorrow you.came
'They told me that tomorrow you would come'
- b. m annə ðəttə ka u kɛʃə tsə l annə
tətə i ɣwajjɛunə
to.me they.have said that the cheese Refl it they.have
taken the boys
'They told me that the cheese was taken by the boys'
- c. m annə ðəttə ka u kɛʃə ts annə
tətə i ɣwajjɛunə
to.me they.have said that the cheese Refl they.have
taken the boys
'They told me that it was the cheese that the boys took'
- d. vujjə ka/kə purə tɛu mənəssə
I.want that also you come(subj)
'I want you too to come'
- e. vujjə ka/kə krɛ purə lɔrə mə'nəssənə
I.want that tomorrow also they come(subj)
'I want them as well to come tomorrow'

(10) *Paulilätino*

- a. m anta nau ka ɣraza b'benizi
to.me they.have said that tomorrow you.come
'They told me that you will come tomorrow'
- a'. m anta nau ka u libru ɖ.aza leddzju
to.me they.have said that the book it you.have read
'They told me that the book you read'
- b. ɛstɛ mməndzɯs ki fintsaza juanni 'bendzɛðɛ kraza
it.is better that even John comes tomorrow
'It is better that John as well comes tomorrow'

(11) *Làconi* (Sardinia)

- a. dɛɔ krɛɔ ka/tɔi issu βuru/ kraza 'eniði
I believe that he too/ tomorrow he.comes
'I believe that he will come as well/tomorrow'
- b. dɛɔ krɛɔ tʃi/ka issu βuru/ kraza 'ɛŋɖʒaða
I believe that he too/ tomorrow he.comes
'I believe that he will come as well/tomorrow'

Let us now consider Ledgeway's (2003a, 2005) analysis, which, as we have seen, consists of a more conservative assumption, namely that complementizer choice depends on mood, and an innovative proposal, namely that movement of CA results in it being spelled out as CHE. The first part of the proposal is in essence the same as that reviewed above for complementizer deletion. Thus our generalization that the 'subjunctive' complementizer is the one that overlaps

with a *wh*-element (if such an overlap is present in the system), remains as inaccessible to Ledgeway as it does to Giorgi and Pianesi and Poletto, and for much the same reason. Indeed, connecting complementizers directly to modality (as is natural in a treatment where both belong to the functional spine of the verb) leaves us without any basis for associating the complementizer with the argumental, nominal system of *wh*-phrases.

The novel part of Ledgeway's (2003a, 2005) proposal is that movement of CA from the Fin to the Force position results in its morphological spell-out as CHE. Now, there is no syntactic and/or morphological theory that can literally change a lexical item into another (because of Chomsky's (1995) Inclusiveness, among others). Nor is it possible to have a lexical entry specified for a context defined by movement (since only intrinsic properties of lexical items belong in the lexicon). Therefore, Ledgeway's analysis seems to imply that double-complementizer languages have a spell-out for indicative complementizers in Fin – and another spell-out for subjunctive complementizers as well as for complementizers in Force. This means that the unification achieved by Ledgeway's analysis is only apparent – as it requires a disjunctive lexical entry for the CHE complementizer. Alternatively, one could try to unify the entry for CHE by recourse to underspecification – i.e. to the idea that CHE is simply the default complementizer of the relevant languages, inserted as the Elsewhere case. But if so, the generalization noted here concerning its overlapping with the *wh*-system would become impossible to explain, in the absence even of modal properties on the complementizer.

2.1.2 Generalized *wh*-complementizers

Section 2.1.1 discusses systems with a split between definite and indefinite complementizers of the type of *Guglionesi* in (1) or in *Paulilätino* in (5). A different grammar, however, seems to be exemplified by *Montenerodomo* in (2) or by *Arena* in (3), in which the *kə/ki* complementizer overlapping with a *wh*-question word is restricted to a few contexts, which in our data coincide with complements to 'after' and 'before'. These contexts do not correlate with the verb modality – which can be subjunctive under 'before' but is consistently indicative under 'after'. Interestingly, all of the languages tabulated in (A') of the Appendix, in which there is no overlap between the *wh*-system and the complementizer system, show the definite vs. indefinite complementizer split of the type in section 2.1.1. By contrast, the distribution of *kə/ki* complementizers in languages like *Arena* in (3) or *Montenerodomo* in (2) appears to be necessarily connected with their *wh*-nature.

In our view the key to the distribution observed resides in the fact that prepositions like ‘before’ or ‘after’ can easily embed complements introduced by *wh*-arguments/adjuncts, as in the comparative structures in Italian (12). In turn, the comparatives in (12) appear to be interpreted in relation to a temporal argument. Thus (12a) corresponds to an LF roughly of the form ‘I arrived at a time *t* before/earlier than a time *t*¹ such that I thought *I would arrive at t*¹’, where the most embedded (italicized) proposition undergoes ellipsis.⁵

- (12) a. Sono arrivato prima di quanto pensassi
 I.am arrived before of how.much I.thought
 ‘I arrived earlier than I thought’
 b. Sono arrivato dopo di quanto stabilito
 I.am arrived after than how.much planned
 ‘I arrived later than planned’

On the basis of (12) we propose that languages that embed a *wh*-complementizer under ‘after’ and ‘before’ effectively select a (comparative-like) *wh*-structure. As noted before, the evidence we are considering is essentially morphosyntactic in nature and therefore hardly insightful when it comes to the semantics; hence semantic analyses are sketched here essentially for the sake of falsifiability. We suggest that the *wh*-complementizer acts in this case as a sort of generalized quantifier, binding the propositional variable, restricted by the following sentence and, at the same time, the temporal variable. We then expect that the generalized quantifier that we have hypothesized also turns up as a *wh* binder of arguments; in other words that it is not specialized as a propositional introducer, but will overlap with *wh*-arguments – as is indeed the case.

It is worth noting that Ledgeway (2009) notices the existence of double-complementizer systems that cannot be assimilated to those in section 2.1.1 in relation to the Calabrian variety of *Cosenza*. He reports that even in the speech of younger generations *chi* is obligatory in optative expressions such as *chi ti vò affucà* ‘(that you) go drown’; in a more conservative variety he finds *chi* to occur in complement sentences when the embedded verb is in the subjunctive (counterfactual). His generalization is that ‘*ca* is a *passé-partout* complementizer compatible both with indicative and with subjunctive, while *chi* is a modally marked complementizer’. Needless to say, this characterization cannot be extended to varieties like *Arena* in (3), in which *ki* is selected by ‘after’.⁶

Let us nevertheless consider his analysis – namely that *ca* realizes the Force head while *chi* lexicalizes the Fin head in Rizzi’s (1997) schema, i.e. the head ‘responsible for marking modal distinctions’. In this schema, therefore, the *Cosenza* variety would involve the opposite distribution to the varieties of section 2.1.1, in which *che* is assigned to the Force position, while *ca* is restricted to Fin. Now, the point being made here is that the *kə* of languages like *Guglionesi*

and the *ki* of languages like *Arena* have in common the fact that they overlap with the *wh*-system. From this perspective, Ledgeway's (2003a, 2005, 2009) analysis would be saying that in double-complementizer systems, the *wh*-complementizer is sometimes a specialized *Fin* and sometimes a default. It seems to us that a theory that captured the continuity between the various double-complementizer varieties would have a clear explanatory advantage.⁷

This leads us to the next point to be examined, namely whether our theory can in fact capture this continuity. Crucially, in all languages of the same type as *Guglionesi* in (1), the indefinite complementizer, in the sense of (6b), turns up in 'before'/'after' contexts as well. Now, if the *wh*-complementizer in languages like *Arena* in (3) or *Montenerodomo* in (2) is to be able to function as some sort of generalized binder of propositional and *wh*-variables, it cannot be definite (which would exclude *wh*-quantification). Therefore, in languages with definite vs. indefinite complementizers like *Guglionesi*, the contexts relevant for its lexicalization will be picked up by the indefinite complementizer, acting as a generalized *wh*-quantifier. This conclusion only depends on the indefinite nature of the complementizer and not on its actual formal identity with a *wh*-argument. Therefore we predict that the indefinite complementizer will also pick up 'before'/'after' contexts in systems where it does not overlap with a *wh*-argument, such as *Paulilätino* in (5).⁸

The final fact worth bringing up is that in all of the languages tabulated in (A–A') in the Appendix, what we have called the indefinite and generalized *wh*-complementizers also introduce relative clauses. Relevant examples are given in (13)–(16). These involve *that* relatives, and not *who* relatives, since, although the relevant complementizers overlap with 'what', as illustrated in (1)–(3) and (5), 'who' has a distinct lexicalization, as illustrated in (13b) and in (14c)–(16c). We can extend to (13)–(16) the analysis that we have already proposed for (12), treating *ki/kə* as some sort of generalized binder for the propositional variable and the argumental one (actually a resumptive pronoun in the examples in (14b)–(16b)). The treatment of 'after' and 'before' complements in what precedes also implies that we should find the indefinite/generalized *wh*-complementizer introducing comparatives. The data in *Paulilätino*'s (16d) confirm this prediction.

(13) *Guglionesi*

- a. ϵ *kkullə kə vvadə səmbrə*
 he.is that that I.see always
 'He is the one that I see all the time'
- b. *ki* *vɛ*
 who comes
 'Who comes?'

(14) *Montenerodomo*

- a. e kkwillə kə vve:tə səmbrə
it.is that that I.see always
'It is the one I see all the time'
- b. e kkwillə kə i so datə lə soldə
it.is that that to.him I.am given the money
'It is the one I gave the money to'
- c. ki ve
who comes
'Who comes?'

(15) *Arena*

- a. su kkijji ki vviju səmpri
they.are those that I.see always
'They are the ones that I see all the time'
- b. su kkijji ki ɲtʃi dəttsi li sərɔdi
they.are those that there I.gave the money
'They are the ones I gave the money to'
- c. ku vɛni
who comes
'Who comes?'

(16) *Paulilätino*

- a. ɛs kussu ki bbio 'zɛmpere
he.is that that I.see always
'He is the one that I see all the time'
- b. ɛs kussu ki dɔd appɔ jau z inari
he.is that that him I.have given the money
'He is the one that I gave the money to'
- a'. juanni ki ɛst ammiɣu meu
John that is friend mine
'John, who is my friend ...'
- c. kie b'beniði
who comes
'Who comes?'
- d. tendzɔ pru llibrɔs ki (nɔ) inarɛ
I.have more books than not money
'I have more books than money'

In a language like *Ardaùli* in (17), of the *Paulilätino* general type, the indefinite complementizer *ki* introduces restrictive relatives and the definite *ka* complementizer introduces appositive relatives. We suggest that in this language the choice of the complementizer is sensitive to the nature of the embedded variable. In particular, *ki* is incompatible with appositive relatives, in that they contain an individual variable; thus, the language resorts to the definite complementizer *ka*.⁹

(17) *Ardaùli* (Sardinia)

- a. *es kussu ki bbio 'sempere*
 he.is that that I.see always
 ‘He is the one that I see all the time’
- b. *es kussu ki dđ appo jau zu inare*
 he.is that that him I.have given the money
 ‘He is the one that I gave the money to’
- a’. *dzuanni ka esti ammiyu meu*
 John, that is friend mine
 ‘John, who is my friend ...’
- c. *kie 'eniđi*
 who comes
 ‘Who comes?’
- d. *tendzo pru llibros ki (no) inare*
 I.have more books than not money
 ‘I have more books than money’

As discussed in [chapter 1](#), Arsenijevic (2009) and Kayne (2010) propose theories of complementation as relativization. We may wonder how these fare with respect to the relativization facts in (13)–(17). Consider for instance *Paulilätino* – i.e. a language that has a two-complementizer system, but in which relative clauses are formed without exception by the indefinite complementizer. If the definite *ka* complementizer is nothing but a relative pronoun, why doesn’t it ever overtly appear in relatives? In fact, the preliminary question arises of how Kayne or Arsenijevic would deal with double-complementizer systems. In Arsenijevic’s terms, different complementizers could select for different types of ‘Force’ arguments. But even so, it is difficult to see why what we have characterized as the indefinite complementizer has the same form as the relative complementizer, but the definite one does not. For it does not seem possible to say that relative clauses associate with a particular type of Force. Kayne’s proposal would have to deal with the facts in a similar way; therefore the same critique applies (cf. also the discussion on conditionals in the next section).

2.2 ‘If’

The complementizer system of Italian, as discussed in [chapter 1](#), and that of the varieties considered so far, have a specialized lexicalization for the interrogative/hypothetical complementizer, *se* in Italian, corresponding (more or less) to English ‘if’. Thus the Italian data in (1)–(4) of [chapter 1](#) are completed by (18).

- (18) a. Se piove (spesso) esco
 if it.rains often I.go.out
 ‘If it rains I (often) go out’
 b. Non so se viene
 not I.know if he.comes
 ‘I don’t know if he’ll come’

As it turns out, this state of affairs does not hold of necessity, but only as one of several possible parametric options. In some Sardinian varieties with a double declarative complementizer system, the indefinite complementizer also lexicalizes hypothetical and interrogative ‘if’, as exemplified by *Làconi* in (19). Note that the indefinite complementizer *tʃi* belongs to the *wh*-lexical series, exemplified by *tʃinni* ‘who’ in (19ci). Note also that the same distinction holds between the complementizer for restrictive and appositive relatives as we discussed for *Ardaùli* in (17); as before, it is the complementizer introducing restrictive relatives that overlaps with the indefinite/‘if’ complementizer.

(19) *Làconi*

- a. m anti nau ka ennis kraza
 to.me they.have said that you.come tomorrow
 ‘They told me that you come tomorrow’
 b. i. bɔʒɔ tʃi ɛɲɔʒas kraza
 I.want that you.come tomorrow
 ‘I want you to come tomorrow’
 b. ii ε m'mɛɲɔʒuzu tʃi b'βɛɲɔʒaza
 it.is better that you.come
 ‘It is better for you to come’
 b. iii sɛɔ essia prima ʒi frssaz arribbau
 I.am gone.out before that you.were come
 ‘I went out before you came’
 c. i. a tʃinni 'tserriaza
 to whom you.call
 ‘Who are you calling?’
 c. ii. itta faizi
 what you.do
 ‘What are you doing?’
 d. i. funti 'gussuzu tʃi tserriu zɛmpɛ
 they.were those that I.call always
 ‘They were the ones that I always call’
 d. ii sɔrrɛ ðua ka ɛstɛ ammiya mia
 sister yours that is friend mine
 ‘Your sister, who is a friend of mine ...’
 e. i tʃi 'ɔllɛzɛ βɛɲɔʒɔ
 that you.want I.come
 ‘If you want, I will come’

- e. ii tʃi ʼβræðɛ nɔ bbessɔ
 that it.rains not I.go.out
 'If it rains, I won't go out'
- f. nɔ iʃʲiu tʃi dɖu tserriu
 not I.know that him I.call
 'I don't know if I shall call him'

Another major pattern of lexicalization of 'that' and 'if' is exemplified in (20) by *Miglionico*. This has a single declarative complementizer and a separate complementizer for hypotheticals and questions – reproducing the conditions of a language like Italian. However, the hypothetical/question complementizer actually belongs to the *wh*-system, in fact overlapping with the *wh*-element for 'who', namely *tʃi*. Because of this, the embedded sentence in (20e) is ambiguous between the *wh*-question reading 'who' and the yes-no question reading 'if'.

(20) *Miglionico* (Lucania)

- a. mə vɔnnə dittə ka viənə kra
 to.me they.have said that he.comes tomorrow
 'They told me that he comes tomorrow'
- b. sɔ assutə primə ka tu və'niəssə
 I.am gone.out before that you came
 'I went out before you came'
- c. i. tʃi ve:nə
 who comes
 'Who is coming?'
- c. ii. tʃe fʲa ʃənə
 what they.do
 'What are they doing?'
- d. tʃi vʲvuo vɛɲnə
 if you.want I.come
 'If you want, I come'
- e. nan sattʃə tʃi venə
 not I.know who/if he.comes
 'I don't know who comes/if he comes'

In short, in the *Làconi* examples in (19) there is a single lexicalization for the indefinite declarative complementizer and for the question/hypothetical complementizer. Table (B) of the Appendix summarizes varieties for which this distribution holds; all of them are Sardinian, of a type in which there is no overlap between the complementizer and the *wh*-system. Crucially, from the table in (B) it can be seen that the hypothetical/interrogative complementizer never has the same form as the *ka* definite complementizer. In turn, in a language like *Miglionico* in (20) the lexicalization for the hypothetical/interrogative

complementizer overlaps with that of a *wh*-element, while the (single) declarative complementizer has a different lexicalization. The data summarized in table (C) of the Appendix show that it is always the hypothetical complementizer that coincides with the *wh*-system (as opposed to the declarative one).

The complementizer system of languages like (19)–(20) closes a potential gap in the discussion so far. As discussed in [chapter 1](#), we know that *wh*-items are not intrinsically interrogative; in Italian, for instance, argument *che* can head a free relative as in (21a) or an exclamative as in (21b). We interpreted these data as showing that question, declarative or exclamative values are not intrinsic to the *wh*-element, but rather contributed by contextual operators. This is far from an isolated case in natural languages. For instance, so-called N-words in Romance are not intrinsically negative, but rather introduce a variable which can receive a negative interpretation in the scope of a negative operator – and can equally well be licensed in non-negative contexts (cf. [chapters 3–4](#)).

- (21) a. Fai che ti pare
do what you please
'Do as you please'
- b. Che non farebbe!
what not he.would.do
'What wouldn't he do!'

On these grounds we expect that a complementizer belonging to the *wh*-series will not require a question interpretation. What does represent a potential problem for the present theory, however, is that none of the complementizers considered in [section 2.1](#) allows for such an interpretation. Nothing in what we have said so far blocks the possibility that a *k*-complementizer could be interpreted as interrogative in the scope of an appropriate question operator, in a similar way to specialized interrogative *wh*-complementizers like English *whether*. From this perspective, systems like (19)–(20) are interesting in the first instance because they instantiate precisely the possibility we predict.

Let us consider, then, what the characterization of the hypothetical/interrogative complementizer may be, beginning with a language like Italian (or indeed English) in which it has a specialized lexicalization. Hypothetical sentences are interpreted much in the same way proposed previously for complement sentences, that is, as the restriction of a propositional variable. According to Lewis's (1975) classical treatment, the latter is bound by an adverb of quantification, or by a generic quantifier in the absence of other overt quantifications. Thus the sentence in (18a) has an LF of the type in (22a), if the adverb of

quantification is computed. Otherwise the propositional variable is closed by a universal quantification, as in (22b).

- (22) a. For many situations/possible worlds x : x it rains, I go out (in x)
 b. For all situations/possible worlds x : x it rains, I go out (in x)

An element which introduces a propositional variable (restricted by the embedded propositional content) is associated in the present grammar with structures of the type argued for in [chapter 1](#) for k -type complementizers – which we will then extend to Italian *se* ‘if’.¹⁰ The latter can in turn be interpreted not only in a hypothetical context as in (15), but also in an interrogative context – exactly like English ‘if’. The embedded interrogative interpretation simply follows if the variable introduced by *se* and the like (and restricted by the proposition following it) can be licensed in the scope of a question operator. We provisionally assume that this is what the interpretation of an interrogative complementizer amounts to, as in (23).¹¹

- (23) I wonder for which situations/possible worlds x : x John comes

At this point the characterization of the *se* complementizer is fundamentally the same as for the *che* complementizer, and more specifically for the indefinite complementizer in (6b), where the variable that the complementizer introduces is closed quantificationally. Therefore we fully expect that ‘if’ and the indefinite declarative complementizer can be lexicalized by the same item, yielding the *Làconi* system in (19).

In view of the proposals reviewed in [section 2.1.2](#) to the effect that complement sentences are relatives, it is particularly worth remarking on the independent literature treating conditionals as relatives – specifically as free relatives. Thus, for Bhatt and Pancheva (2006), ‘turning to conditionals, our proposal that they are interpreted as free relatives amounts to the claim that they are definite descriptions of possible worlds’. As for the syntax, ‘the null operator in Spec, CP of *if*-clauses and likely the *when* itself in e.g. German conditionals, is a definite binder of the possible world variable’. As Bhatt and Pancheva note, the proposal that there is a covert operator in the Spec, CP of conditional *if* clauses goes back to Larson (1985), where the focus of the discussion is interrogative *if* clauses. For the latter, Larson posits the presence of a covert *whether*, which he extends to conditional *if* clauses. For Bhatt and Pancheva, ‘the fact that *if* functions in many languages as both a conditional and an interrogative complementizer makes sense within the general proposal that conditionals are free relative clauses. In English and in many other languages, this syncretism would be part of a more general structural parallelism between questions and free

relatives'. However, if 'free relatives are interpreted as definite descriptions, i.e. with the variable abstracted over being bound by a definite operator' (Bhatt and Pancheva 2006), the lexicalization patterns observed in *Làconi*, in which 'if' overlaps with the indefinite complementizer, and not with the definite one, seem to be difficult to account for. More generally, the present account does not require us to postulate zero world operators in the syntax. Quite simply, the so-called complementizer is itself the element that introduces a (situation/possible world) variable in the syntax of LF.¹²

If anything, our problem is how to provide a sharper characterization for languages like Italian that have two distinct complementizers, or languages like *Paulilàtino* in (5) that end up having three. Let us begin by considering a language with three distinct forms, including the definite and indefinite declarative complementizers, as well as the hypothetical/interrogative complementizer, as exemplified for *Paulilàtino* in (5) and (24). Following the discussion in section 2.1, the definite complementizer introduces a propositional complement akin to a definite description, while the indefinite complementizer introduces an indefinite variable. The latter is also true of the proposition introduced by 'if', which therefore needs to be further differentiated.

- (24) *Paulilàtino*
 sì 'prɔɛðɛ nɔ ɛssɔ
 if it.rains not I.go.out
 'If it rains, I won't go out'

In introducing the contexts triggering subjunctive in Italian in section 2.1.1, we mentioned that they are not necessarily lexically selected; rather, subjunctive can be triggered in the scope of the question operator in (8c). In the scope of the question operator, the declarative complementizer, ordinarily embedded by a verb like *to know*, alternates with the 'if' complementizer, as in (25) – a phenomenon discussed in the literature under the name of 'unselected questions' (Adger and Quer 2001). Needless to say, the reading of the embedded sentence in (25) is always non-factive, i.e. within the scope of the matrix question operator, independently of whether the indicative or the subjunctive is selected. As shown in (26), a matrix negation has very much the same effect as a question operator, both with respect to the triggering of the subjunctive in (26a) and to the possibility of the 'if' complementizer in (26b). In (26) only the co-occurrence of the *che* complementizer with the embedded indicative in (26a) yields a presupposed (factive) reading of the embedded sentence; both the subjunctive and the 'if' complementizer force the sentence to be read within the scope of the negation operator.

- (25) Sai se è/ sia venuto?
 you.know if he.is/ be come
 'Do you know if he has arrived?'
- (26) a. Non sa che sono/sia guarito
 not he.knows that I.am healed
 'He doesn't know that I am healthy again'
- b. Non sa se sono/sia guarito
 not he.knows that/if he.be healed
 'I don't know if he is healthy again'

The fact that 'if' is triggered in the scope of a matrix question or negative operator with an otherwise declarative (factive) verb, as in (25)–(26), points the way to a solution for the problem we are concerned with, i.e. the characterization of the contexts specialized for 'if'. Negations and questions are two of the fundamental contexts triggering (negative) polarity items – the third one being the hypothetical one. Thus, for instance, in Italian *nessuno* 'anybody', *niente* 'anything' etc. are licensed in the scope of negation (yielding the equivalent of the negative quantifiers *nobody*, *nothing* etc. in English), as well as in the scope of a question or of a hypothetical. We suggest that the complementizer system of languages like Italian or *Paulilätino* in (24) (or, for that matter, English) is sensitive to polarity, so that complementizers, i.e. nominal heads responsible for introducing propositional variables, are lexicalized by two different items in non-polarity and polarity contexts, surfacing in the latter as *se* in Italian, *if* in English etc. Correspondingly, we shall refer to these as the polarity complementizers.

In section 2.1 we concluded that it is to be expected that the complementizer system, i.e. the system of propositional variable introducers, and the *wh*-system may coincide. More to the point, in cases of split complementizer systems the coincidence is predicted to involve the indefinite complementizer, i.e. the one introducing an indefinite variable, rather than the definite complementizer, which introduces a (propositional) definite description. This mode of reasoning can now be extended to the 'if' complementizer. In languages of the type of *Làconi* in (19) there is no overlap between the complementizer and *wh*-systems. However, languages like *Miglionico* in (20) verify this prediction for the systems with a polarity complementizer 'if' and a non-polarity complementizer 'that'. For in *Miglionico*, as in the other varieties tabulated in (C) of the Appendix, it is the 'if' complementizer that belongs to the *wh*-system – and not the (single) declarative complementizer.

A further parameter brought to the fore by varieties like *Miglionico* has to do with which element of the *wh*-system the complementizer identifies with. In

Italian, as reviewed in [chapter 1](#), the overlap is between the declarative complementizer and the *wh*-element for ‘what’. By contrast, in *Miglionico* in (20) the overlap is between the ‘if’ complementizer and the *wh*-item for ‘who’. This parameter is independent of the others considered. Thus, in *Ruvo di Puglia* in (27) the ‘if’ complementizer has the same form as ‘what’, and not ‘who’ as in *Miglionico*. By contrast, in a variety like *Secinaro*, otherwise entirely comparable to *Guglionesi* in (1), the modal complementizer has the same form as ‘who’, and not ‘what’, as in (28).

(27) *Ruvo di Puglia* (Apulia)

- a. tʃə cɔ:və nən iəssə
that it.rains not I.go.out
‘If it rains I won’t go out’
- a’. nən tsattʃə tʃə frattə mə cɔmə
not I.know if brother-yours me calls
‘I don’t know if your brother is calling me’
- b. tʃi vənə
who comes
‘Who comes?’
- c. tʃe ffəʃə
what you.do
‘What are you doing?’

(28) *Secinaro* (Abruzzi)

- a. vujjə kə vvi
I.want that you.come
‘I want you to come’
- b. kə vve
who comes
‘Who comes?’
- c. ku ffə
what you.do
‘What are you doing?’

In terms of the discussion in [chapter 1](#), the *wh*-quantifier for ‘what’, such as Italian *che*, is characterized by the absence of any lexical restriction. Thus *che* questions can be answered by an animate or inanimate noun phrase or by a proposition, as in (29). Of course a lexical restriction can be added by a noun phrase, when *che* ‘what’ appears as its specifier – or by a sentence, when *che* appears as the ‘that’ complementizer.

- (29) a. Che c’ è?
what there is
‘What is it?’

- b. Il campanello
 'the doorbell'
 b'. Il postino
 'the postman'
 b''. Suona il postino
 rings the postman
 'The postman is ringing'

By contrast, the *wh*-quantifier for 'who' is characterized by a restriction to humans, which poses an interesting question as to its ability to appear as a complementizer at all. In fact, the occurrence of, say, *Miglionico tʃi* in (20) as a complementizer excludes the possibility that the restriction to humans is encoded in the lexicon. What we suggest instead is that it is a lexical property of *tʃi* that it requires a restriction. In complementizer contexts the latter is provided by the proposition that *tʃi* introduces. In *wh*-contexts, on the other hand, the restriction to humans may represent an interpretive closure in the absence of lexical restrictions.

If this line of reasoning is correct, we should be able to find languages in which the *wh*-quantifier for 'who' also appears as the *wh*-determiner, restricted by a following noun phrase. A case in point is *Fontanigorda* in (30), which, like Italian, has a single declarative complementizer overlapping with a *wh*-quantifier – except that the overlap involves 'who', rather than 'what'. As expected, the same element also appears as the *wh*-determiner, as shown in (30d).

(30) *Fontanigorda* (Liguria)

- a. m aŋ ittu ke te vie dopu
 to.me they.have said that you come afterwards
 'They told me that you are coming afterwards'
- b. ke te tʃammi
 who you call
 'Who do you call?'
- c. kuɔse te fe
 what you do
 'What are you doing?'
- d. ke kamiʒa te te betti
 which shirt you yourself put.on
 'Which shirt are you putting on?'

2.3 The interaction with (non-)finiteness

The last question we shall consider in this chapter concerns yet another asymmetry between *wh*-quantifiers and complementizers, observed even in

languages in which they have the same lexicalization. Thus, in standard Italian *che* ‘what’ introduces both finite and infinitival questions, as in (31a). In contrast, complementizer *che* is restricted to finite contexts, and infinitival embeddings involve, if anything, prepositional introducers like *di* ‘of’ in (31b).

- (31) a. Che faccio/ fare?
 what I.do/ to.do
 ‘What should one/I do?’
 b. Mi hanno detto di/*che essere venuti
 to.me they.have said to/that be come
 ‘They told me that they had come’

The problem is not a descriptive one, since complementizer *che* can be distinguished from *wh-che* on the basis of both the syntactic context of insertions and its semantic characterization. From an explanatory point of view, however, the objection may legitimately be raised that the connection between complementizers and finiteness crucially links them to the functional projections of the verb. Now, finiteness has two components – one relating to tense, and another relating to agreement. As is well known, some Romance languages allow these two components to be distinguished, since they feature so-called inflected infinitives.

A case in point is provided by Sardinian varieties, including *Paulilätino* in (32). Both inflected and non-inflected infinitives can be introduced by a prepositional complementizer, as in (32b) – but inflected infinitives also allow for the finite complementizer, as independently observed by Jones (1993). The finite complementizer, then, is sensitive not to the temporal/modal/aspectual properties of the verb, but to the presence of an agreement inflection. As we have seen, *Paulilätino*, like many Sardinian varieties, has two declarative complementizers, a modal and a non-modal one – and it is the modal complementizer that co-occurs with inflected infinitives. However, Manzini and Savoia (2005) also provide examples from varieties which have only one declarative complementizer, which also appears in front of the inflected infinitive (for instance, *Siniscola*).

- (32) *Paulilätino*
 a. I appu attu innantis de/ ki θorrare- s tuε
 it I.have done before to/ that come.back-2sg you
 ‘I did it before you came back’
 b. I an fattu innantis de 'ennere- (nε) /ki 'ennere-nε 'iççəçə
 it they. have done before to come- 3pl that come-3pl they
 ‘They did it before they came’

In our theory (cf. in particular [chapter 5](#)) the so-called agreement inflection is a lexicalization of the EPP argument of the sentence internal to the verb. If

no lexicalization of the EPP argument is present, either by an inflection or by a lexical subject, as is the case for non-inflected infinitives, we assume that the EPP argument is introduced as a variable at the interpretive interface. This variable can then be given a bound reading (control/raising) or a generic reading (arbitrary control). In such terms, the restriction of the *che*-type complementizer to finite sentences – including inflected infinitivals – can be restated as an incompatibility between it and the EPP variable. The question then becomes why the finite complementizers would have this crucial property. We may surmise that the presence of an EPP variable within the sentence defines an open predicate, rather than a proposition. Since we have proposed that the finite complementizer of Romance, and presumably that of English as well, takes a proposition as its complement, we may conclude that what the complementizer is actually incompatible with is the open expression resulting from the presence of the EPP variable.

Similarly, for Roussou (2010), the fact that in English *that* requires a propositional complement ultimately yields the *that*-t filter, precisely because it excludes the possibility that the EPP argument is a variable. In Roussou's terms, the absence both of *that* and of the EPP gives rise to a case of predicate embedding – which allows the base-generated *wh*-phrase in the matrix left periphery to bind a variable in the argument structure of the embedded predicate. In fact, Roussou (2010) analyses both the **for-t* and the **for-to* filters as results of the same restriction, to the effect that *for*, like *that*, embeds a proposition, barring EPP variables.

Though a discussion of the *that*-t filter is beyond the scope of the present work, it is worth recalling that, according to the classical conclusion of Taraldsen (1978), languages like Italian lack *that*-t filter effects (whence the apparent lack of correlation with the finiteness constraint on the complementizer) in that they are null-subject languages. We propose that what is at stake is simply the fact that while in non-null-subject languages the EPP must be satisfied by a syntactic-level argument (the subject), in null-subject languages a morphological-level argument (the inflection) suffices. The latter therefore closes off the proposition even in the case of *wh*-extraction (effectively acting as a resumptive pronoun). Similarly, non-null-subject Romance languages, in contrast to English, generally have subject clitics and resolve the propositionality requirement on *che*-type complementizers by lexicalizing one of them. On the basis of a comparison with Romansh varieties, Taraldsen (2002) and Manzini and Savoia (2005) in particular conclude that the *-i* of French *qui* is a subject clitic of this type.

A further twist on the finiteness problem is introduced by the 'if' complementizer. In a language like standard Italian, hypothetical *se* 'if' is restricted to

finite sentences, as in (33a), while interrogative *se* normally occurs in front of infinitival questions, as in (33b). Now, recall that in *Làconi* in (19) the same *tʃi* form lexicalizes both the modal declarative complementizer and the polarity (i.e. hypothetical/question) one. As it turns out, while declarative *tʃi* is restricted to finite sentences, interrogative *tʃi* combines with infinitivals, as in (34). In the discussion surrounding (23), we interpreted the interrogative complementizer as a *wh*-element ranging over propositions in the scope of a question operator. We must conclude that this interpretation makes it compatible with embedded EPP variables. In this respect, ‘if’ in the *wh*-construal behaves exactly like other *wh*-elements, including the English *wh*-complementizer *whether*.

- (33) a. *Se piovere, ...
 if to.rain
 b. Non so se uscire
 not I.know if to.go.out
 ‘I don’t know whether to go out’
- (34) *Làconi*
 nɔ iʃʲiu tʃi dɔu tʃɛrri'ai
 not I.know if him to.call
 ‘I don’t know whether to call him’

The issue is further complicated by the fact that while Italian *se* or *Làconi tʃi* can introduce infinitival questions, as in (33)–(34), neither English *if* nor French *si* can do so. In present terms, this means that French *si* and English *if* must embed complete propositions, very much like *che* or *that*, even if they are construed in the scope of an interrogative operator.

The present analysis can be usefully compared with Kayne’s (1991). For Kayne, the finiteness constraint on complementizers reflects their incompatibility with the PRO subject of infinitivals; in other words, he substantially agrees with the view advocated here that the finiteness requirement relates to the EPP rather than to the temporal properties of the sentence. However, for him, what is responsible for this constraint is the structural government relation holding between the complementizer and PRO. The lack of such a constraint on Italian *se* correlates with a further structural property, i.e. with the relatively high position of the infinitive, which protects PRO from government by the complementizer. That the position of the infinitive is higher in Italian than in French is shown, according to Kayne, by the fact that the clitic precedes the infinitive in French, while it follows the infinitive in Italian. However, he acknowledges that his theory ends up not accounting for the incompatibility of Italian *che* with the infinitive, since on the evidence of cliticization the infinitive must be in the same high position in declaratives as in questions (1991: 95). In *Làconi*

the problem is even starker, since the same *tʃi* complementizer cannot combine with infinitival declaratives – as predicted by Kayne on the basis of the preverbal position of the clitic seen in (19) – but can combine with infinitival questions.

Furthermore, in the account of Kayne (1991), *whether* is oblivious to the finiteness of the sentence it embeds because it is not a complementizer at all, but a *wh*-phrase – where these two notions are given a structural characterization once more, as a C head and a C Specifier respectively. However, a structural characterization as a C Specifier cannot be applied to *Làconi tʃi* – on pain of not being able to predict that (as a C head) it excludes non-finite sentences in declarative contexts.¹³

A different type of evidence concerning definiteness restrictions on complementizers of the *k*-series comes from Northern Italian varieties (including some that display the standard Italian identity of form between the *wh*-quantifier for ‘what’ and the complementizer), which form questions by combining *wh* quantifier and complementizer, as discussed in detail in chapter 1 (especially section 1.2). Most of the relevant languages alternate between *wh*-*che* in finite sentences and the *wh*-element alone in infinitivals, as shown in (35) for *Zoldo Alto*. However, there are (a few) otherwise entirely comparable languages in which the *wh*-*che* cluster introduces both finite and infinitival questions, like *Civate* in (36) – though the *che*-type complementizer is otherwise restricted to finite complements.

(35) *Zoldo Alto* (Veneto)

- a. di-me ke ke te faθe
 tell-me what that you do
 ‘Tell me what you are doing’
- b. no sai ke fa
 not I.know what to.do
 ‘I don’t know what to do’

(36) *Civate* (Lombardy)

- a. tʃe ke te tʃamet
 who that you call
 ‘Who are you calling?’
- b. so mia tʃe ke tʃa'ma
 I.know not who that to.call
 ‘I don’t know who to call’

The pattern of *Zoldo Alto* is what we expect on the basis of the assumption that in the *wh*-*che* sequence, the *che*-type element is the ordinary complementizer; as such we predict it to be sensitive to the finiteness of the embedded clause and in particular to exclude an infinitival sentence, whose EPP

argument is a variable under present assumptions. Let us consider, then, the case of *Civate*. Evidently the presence of the *wh*-quantifier in the left periphery of the *ke* complementizer is sufficient for the latter to also be interpreted in the scope of a question operator – that is, as a *wh*-element of sorts ranging over propositions. If so, we expect it to have the property that we have independently reviewed for such elements above, that is, that of being insensitive to the finiteness restriction.

Appendix

(A)	<i>Comp</i> <i>def</i>	<i>Comp</i> <i>indef/wh-</i>	<i>rel</i>	<i>who</i>	<i>what</i>
<i>Sonnino</i>	ka	ke	ke	ki	ke
<i>Pontecorvo</i>	ka	kə	kə	ki	kə
<i>Colledimacine</i>	ka	kə	kə	ki	kə
<i>Montenerodomo</i>	ka	kə	kə	ki	kə
<i>Torricella Peligna</i>	ka	kə	kə	kə	kə
<i>Secinaro</i>	ka	kə	kə	kə	ku
<i>Civitaluparella</i>	ka	kə	kə	ki	kə
<i>Vastogirardi</i>	ka	kə	kə	kia	kə
<i>Capracotta</i>	ka	kə	kə	kia	kə
<i>Guardiaregia</i>	ka	kə	kə/ka	ki	kə
<i>Gugliesesi</i>	ka	kə	kə	ki	kə
<i>Frigento</i>	ka	ke	ki	ki	ke
<i>Nocera</i>	ka	kə	kə	kə	kə
<i>Albidona</i>	ka	kə	kə	kwe	kə
<i>Terranova Pollino</i>	ka	kə	kə	ku	kə
<i>Morano</i>	ka	ki	ki	ku	ki
<i>Orsomarso</i>	ka	ki	ki	ku	ki
<i>Conflenti</i>	ka	ki	ki	kine	ki
<i>Platania</i>	ka	ki	ki	kine	ki
<i>Gizzeria</i>	ka	ki	ki/ka	kina	ki
<i>Sorbo S. Basile</i>	ka	ki	ki	kine	ki
<i>S. Pietro a Maida</i>	ka	ki	ki	ku	ki
<i>Iacurso</i>	ka	ki	ki	ku	ki
<i>Arena</i>	ka	ki	ki	ku	ki
<i>Umbriatico</i>	ka	ki	ki	kini	ki
<i>Gerace</i>	ka	ki	ki	ku	ki
<i>S. Agata del Bianco</i>	ka	ki	ki	ku	ki
(A')					
<i>Dorgali</i>	ka	ki	ki	kie	itte
<i>Aritzo</i>	ka	tʃi	tʃi	(e)tʃine	itte
<i>Ardaùli</i>	ka	ki	ki – ka	kie	(e)itte
<i>Paulilàtino</i>	ka	ki	ki	kie	itte
<i>Gallo</i>	ka	ku	ku	ki	kwe

<i>Tufillo</i>	ka	kə	kə	kia	(kə) dɛkkə
<i>Canosa Sannita</i>	ka	kə	kə	ki	ki
(B)	<i>comp</i>	<i>comp/ if</i>	<i>rel</i>	<i>who</i>	<i>what</i>
<i>Làconi</i>	ka	tʃi	tʃi – ka	tʃinni	itta
<i>Orroli</i>	ka	ki	ki – ka	(e)kini	itta
<i>Settimo S. Pietro</i>	ka	ki	ki	kini	itta
<i>Siliqua</i>	ka	ki	ki	(ak)kini	itta
(C)	<i>comp</i>	<i>if</i>	<i>rel</i>	<i>who</i>	<i>what</i>
<i>Miglionico</i>	ka	tʃi	ka	tʃi	tʃe
<i>Minervino Murge</i>	ka	tʃə	ka	tʃə	tʃəkə
<i>Gravina in Puglia</i>	ka	tʃi	ka	tʃi	attʃokə
<i>Bitetto</i>	ka	tʃə	ka	tʃə	tʃə
<i>Ruvo di Puglia</i>	ka	tʃə	ka	tʃi	tʃə
<i>Canosa</i>	ka	tʃə	ka	tʃə	tʃɛ
<i>Bisceglie</i>	kɛ	tʃə	kɛ	tʃə	tʃə
<i>Martina Franca</i>	ka	tʃə	ka	tʃə	tʃə
<i>Molfetta</i>	ka	tʃi	ka	tʃi	tʃɛ
<i>Mesagne</i>	ka	tʃi	ka	tʃi	tʃe
<i>Putignano</i>	ka	tʃə	ka	tʃə	tʃɛ
<i>Brindisi</i>	ka	tʃi	ka	tʃi	tʃɛ
<i>Grottaglie</i>	ka	tʃi	ka	tʃi	tʃe
<i>Carmiano</i>	ka	ʃi	ka	ʃi	tʃɛ
<i>Uggiano</i>	ka	tʃi	ka	tʃi	tʃɛ
<i>Copertino</i>	ka	tʃi	ka	tʃi	tʃe
<i>Melissano</i>	ka	tʃi	ka	tʃi	tʃi

3 *Sentential negation: adverbs*

The currently standard approach to the structure of negation (in Romance) was inaugurated by Pollock (1989), who proposed that negation adverbs such as *pas* in French fill the Spec position of a NegP projection generated below the I position targeted by the verb. The head of NegP can in turn be filled by a negative clitic like *ne* in French, whose higher inflectional position depends on movement, i.e. cliticization. In other languages, which include colloquial French, no negative head is present. Belletti (1990) applies the same theory to Italian, which only has a negative head; the latter originates in the Neg position and moves higher as a result of cliticization.

A more complex set of data, involving Northern Italian varieties, is considered by Zanuttini (1997), who proposes that there are several Neg positions. Specifically, a Neg position is generated above I, while below I there are three Neg positions. The inflectional Neg position hosts negative clitics in languages like Italian which do not require a sentential negation adverb. In contrast, languages which require a sentential negation adverb generate it in one of the lower Neg positions; if a clitic combines with the adverb, it is generated in the head of the relevant Neg position and moves to the inflectional domain via cliticization. In other words, for Zanuttini (1997) preverbal clitic negations are associated with two different structures, according to whether they negate alone or combine with a negative adverb. In turn, the lower Neg positions are defined in relation to the general hierarchy of adverbs proposed by Cinque (1999); according to Zanuttini (1997), her Neg₂, Neg₃ and Neg₄ positions occur within the aspectual adverbial series, while they do not interact in any significant way with either the temporal or the modal series. Neg₁ corresponds to the negative clitic position in the inflectional domain.

According to Cinque (1999: 106), the aspectual adverbial series is ordered according to the hierarchy partially reproduced in (1a). The three adverbial negation positions proposed by Zanuttini (1997: 99) are ordered with respect to this hierarchy as in (1b). The three different Neg positions correspond to three different types of sentential negation adverbs, which Zanuttini individuates in