Doubling in relative clauses

Aspects of morphosyntactic microvariation in Dutch

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: word cloud by www.wordle.net

ISBN: 978-94-6093-100-0
NUR 616

Doubling in relative clauses

Aspects of morphosyntactic microvariation in Dutch

Verdubbeling in relatiefzinnen

Aspecten van morfosyntactische microvariatie

in het Nederlands

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 18 januari 2013
des middags te 4.15 uur

door

Eefje Pietertje Maria Boef

geboren op 21 december 1983 te Utrecht

Promotoren: Prof.dr. L.C.J. Barbiers
Prof.dr. N.F.M. Corver

Voor mijn ouders

Contents

Acknowledgments

First and foremost, I thank my supervisors – Sjef Barbiers and Norbert Corver – for their invaluable advice, guidance and encouragement. Above all, I thank them for having faith in me and for giving me the freedom to go my own way. It has been a pleasure and privilege to work with them.

Even though Hans Bennis, Marika Lekakou and Marcel den Dikken were not officially my supervisors, they deserve to be mentioned here as well. I thank Hans Bennis for his trust and for the opportunities he gave me. I thank Marika Lekakou for always taking the time to discuss my work and for the kind words of support and advice. I thank Marcel den Dikken for his generosity with his time and knowledge, and for invaluable discussion and feedback. His comments largely increased the quality of this thesis.

I thank the members of the committee – Hans Bennis, Lisa Cheng, Marcel den Dikken, Marika Lekakou and Eric Reuland – for reading and approving the manuscript, and I am grateful to all the other people that have contributed to this thesis by taking the time to discuss my work with me, either in real life or via email. In this respect, I warmly thank Hans Broekhuis, Guglielmo Cinque, Leonie Cornips, Lena Dal Pozzo, Marijke De Belder, Karen De Clercq, Jakub Dotlačil, Lizet van Ewijk, Antonio Fábregas, Irene Franco, Bettina Gruber, Mirjam Hachem, Liliane Haegeman, Wilbert Heeringa, Marko Hladnik, Jacqueline van Kampen, Richard Kayne, Olaf Koeneman, Hilda Koopman, Marjo van Koppen, Tom Leu, Andrew Nevins, Marc van Oostendorp, Dennis Ott, Andreas Pankau, Gertjan Postma, Gillian Ramchand, Juan Romeu, Johan Rooryck, Eddy Ruys, Martin Salzmann, Ankelien Schippers, Radek Šimík, Peter Svenonius, Adam Szczegielniak, Tarald Taraldsen, Tanja Temmerman, Gary Thoms, Øystein Vangsnes, Heimir Freyr Viðarsson and Hedde Zeijlstra. The research presented in this thesis has furthermore benefitted from discus-

sion with people I met at conferences, workshops and other events. There are simply too many of them to be named in person.

My first linguistics home in the Netherlands over the past four years was the Meertens Instituut in Amsterdam. I profoundly appreciate the opportunities I have been given there and I thank all my colleagues for providing a stimulating working environment. A very special thanks goes out to Franca Wesseling, with whom I shared an office for almost three years, for her friendship and support. I am happy that she accepted to be my *paranimf*, and I hope we will share a future in linguistics. I furthermore thank Rob Zeeman for technical assistance with the data and Jan Pieter Kunst for helping me with the maps. Finally, I thank the Meertens Panel informants for their participation and patience in filling out two lengthy questionnaires.

My second home was the Utrecht institute of Linguistics OTS (UiL OTS) at Utrecht University, where I finished my studies in linguistics and always felt welcome. For that, I am grateful to all my colleagues at UiL OTS and at the Department of Dutch Studies. In particular, I thank Alexis Dimitriadis for his help with LaTeX, and I warmly thank Bettina Gruber and Mirjam Hachem for their company and friendship.

During my time as a PhD student, I have been fortunate enough to have had the opportunity to travel. In 2010, I spent the spring semester at the Center for Advanced Study in Theoretical Linguistics (CASTL) at the University of Tromsø (supported by an Yggdrasil grant of the Research Council of Norway). I thank all the people at CASTL for their warm welcome and for providing an inspiring and supportive environment, both socially and intellectually. Special thanks go out to my Tromsø friends: Lena Dal Pozzo, Irene Franco, Przemysław Jasiński, Peter Jurgec and Juan Romeu for making my time in the Arctic unforgettable. I am very happy that I still see most of them regularly. In the beginning of 2011, I spent almost three months in New York, at CUNY and NYU. I thank everybody there for welcoming me. Special thanks go out to Andrew Bonfilio, Karen Duek, Christen N. Madsen II and my flatmates at Canal Street for their excellent company. It was *gourmazing*.

I feel blessed for having always been surrounded by amazing friends, old and new. Some of them have already been mentioned above. Here I want to mention Inger van den Bos, Carmen Jacobs, Mieke van der Meij, Jolien Scholten, Jan Pieter Toussaint, Hanneke Vos, Franca Wesseling and Tom Zinger. I thank them for all the great times we spent together (dinners, drinks, awesome nights out, weekend trips and holidays) and for reminding me there is so much more to life than linguistics. I also thank the girls at Oudkerkhof in Utrecht, with whom I shared a great home: Lotte Agterkamp, Noortje van den Berg, Joke Bolt and Birgit Oldenhof.

My final thanks go out to my family for their unconditional love, support and encouragement. I thank Maarten for being the best big brother a girl can ever wish for. I thank Nynke for understanding what it is to do a PhD, for being my *paranimf* and for the great collages; they always made me laugh. I thank Jeske for all the fun we have together and for always having my back. I thank my parents, Nico and Simone, for believing in me and for always being there. *Dank jullie wel!*

Berlin, October 2012

Abbreviations

1	first person
2	second person
3	third person
ACC	accusative case
ANS	*Algemene Nederlandse Spraakkunst* (Dutch reference grammar)
C	common (=non-neuter) gender
CA	complementizer agreement
CLITIC	clitic pronoun
DAT	dative case
DEF	definite
DEM	demonstrative pronoun
DFC	doubly filled COMP
FCC	finite complement clause
FIN	finite
GEN	genitive case
HEA	Head External Analysis
HRA	Head Raising Analysis
INF	infinitival
INST	instrumental
M	masculine gender
MA	Matching Analysis
MPQ	Meertens Panel Questionnaire
N	neuter gender
NOM	nominative case
OBJ	object
Op	operator
Op_Q	question operator
Op_{REL}	relative operator

PAST	past tense
PERF	perfect tense
phi/ϕ-features	person, number and gender features
PL	plural
Q	question
QR	quantifier raising
RC	relative clause
RP	relative pronoun
SAND	*Syntactische Atlas van de Nederlandse Dialecten* (Syntactic Atlas of the Dutch Dialects)
SG	singular
s/o asymmetry	subject/object asymmetry
STR	strong agreement
STRONG	strong pronoun
SUBJ	subject
WEAK	weak pronoun
wh-Q	*wh*-question
WK	weak agreement

CHAPTER 1

Preliminaries

This thesis presents a study of (long-distance) restrictive relative clauses (RCs) and related constructions in Dutch. RCs have always featured prominently in linguistic theory, as a result of which there exist now many different analyses of RCs, but there are as many outstanding questions and unresolved issues. The goal of this thesis is to contribute to the study of RCs and related constructions by looking at previously undiscussed *microvariation*, and more specifically, by looking at *doubling* in Dutch (long-distance) RCs. Doubling constructions shed new light on several outstanding questions within generative linguistics, concerning the structure of A-bar chains (chapter 2), the syntax of RCs, the structure of the left periphery (chapter 3), and the nature of relative pronouns and complementizers (chapter 4).

1.1 Introduction

A *restrictive* relative clause (RC) is a subordinate clause that modifies a noun phrase, the *RC head*.[1,2] Besides having a function in the matrix clause, the RC head has a function inside the RC as well. In (1), for instance, the RC head *man* functions as the object of the matrix verb *to know*, and at the same

[1] I will primarily be concerned with *restrictive* RCs and disregard other types of RCs, because – as extensively discussed in the literature (cf. de Vries 2002 for an overview) – restrictive RCs pattern differently in a number of respects from other types of RCs (although unified accounts of different types of RCs have been proposed, see e.g. Cinque 2010).

[2] The notion *head* in the context of RCs might be a bit confusing. It is not a head as familiar from *X-bar theory* (Chomsky 1970, Jackendoff 1977). Rather, the notion *RC head* can best be defined as the element that is being modified by the RC (the *modifiee*), or as the *antecedent* of the relative pronoun.

time it functions as the object of the verb *to call* inside the RC. It is this *pivot* function of the RC head (cf. De Vries 2002) that has led to much interest in RCs.

(1) I know the man [$_{RC}$ they have called].

Standard Dutch restrictive RCs follow the RC head they modify (i.e. they are *head-initial* or *postnominal*) and they need to be introduced by a relative pronoun (RP).[3] The relativized item leaves a gap (not a resumptive pronoun) at the extraction site inside the RC. This is illustrated in (2) for a Standard Dutch object RC with the 3rd person singular common gender RC head *man* 'man' that requires the common gender relative pronoun *die* 'that'.[4]

(2) Ik ken de man [$_{RC}$ **die** ze _ geroepen hebben].
I know the man RP they called have
'I know the man they have called.' [Standard Dutch]

Dutch also features what I refer to as *long-distance relativization*: a RC construction in which there is an extra embedding and the RC head is related to the gap in the most deeply embedded clause. As illustrated in (3) for a long-distance object RC in Standard Dutch, the RC itself is introduced by the relative pronoun and the lower finite embedded clause is introduced by the declarative complementizer *dat* 'that'.[5]

(3) de man [$_{RC}$ **die** ik denk [**dat** ze _ geroepen hebben]]
the man RP I think that they called have
'the man who I think they have called' [Standard Dutch]

There is a wealth of morphosyntactic variation with respect to this particular construction. Most variation is found in the form and the nature of the elements that introduce the RC itself and that introduce the lower clause, e.g. *d*-pronouns, *w*-pronouns, complementizers or 'doubly filled Comp' configurations. The latter are structures in which a pronoun and a complementizer simultaneously appear in the left periphery (Comp) of an embedded clause.[6]

[3] I will follow common practice and use the term *relative pronoun* descriptively to refer to the pronoun introducing a RC. However, I do not assign any theoretical status to this term, as I do not take there to be such a thing as a *relative pronoun* (at least for Dutch, cf. Wiltschko 1998). In chapter 4, I come back to the issue of the nature and status of relative pronouns.

[4] In this introductory chapter, I simply indicate the gap position inside the RC with _, and remain agnostic about what this is the base position of (the relative pronoun or the relative pronoun *plus* the RC head). See chapter 3 for detailed discussion.

[5] Long-distance extraction in relative clauses as in (3) is accepted by fewer speakers than long-distance extraction in *wh*-questions, i.e. it is generally judged less grammatical. See section 2.3 for details.

[6] The left periphery of the clause was traditionally referred to as the Comp position, hence the notion 'doubly filled Comp'.

In colloquial Dutch, relative pronouns can be *doubled*, in which case a relative pronoun introduces the RC itself as well as the finite embedded clause. This is illustrated in (4) for the relative pronoun *die*.[7]

(4) de man [$_{RC}$ **die** ik denk [**die** ze _ geroepen hebben]]
the man RP I think RP they called have
'the man who I think they have called'
[colloquial Dutch, cf. Barbiers et al. 2005:85]

The main goal of this thesis is to provide a principled account of the attested microvariation in long-distance RCs in general, and the doubling construction in (4) in particular. Chapter 2 shows that doubling of the relative pronoun in colloquial Dutch long-distance RCs like (4) is best analyzed as the result of successive-cyclic movement via SpecCP together with multiple copy spell out. Chapter 3 demonstrates that (this proposed account of) doubling challenges current views on the syntax of RCs and proposes a specific implementation of the traditional Head External Analysis for Dutch RCs. Chapter 4 is concerned with the nature of the elements involved in doubling constructions, namely (relative) pronouns and complementizers, and claims that the more underspecified an element is, the more different syntactic environments it may appear in.

Section 1.2 of this introductory chapter lays out a number of theoretical notions and assumptions that are instrumental to the analyses to be presented and discussed in this thesis. The methodology and design underlying the data collections that form the empirical basis for the studies in this thesis are presented in section 1.3. Finally, section 1.4 provides an outline and outlook of this thesis, in which the most important empirical results and theoretical claims will be briefly summarized.

1.2 Theoretical assumptions

1.2.1 The model of grammar

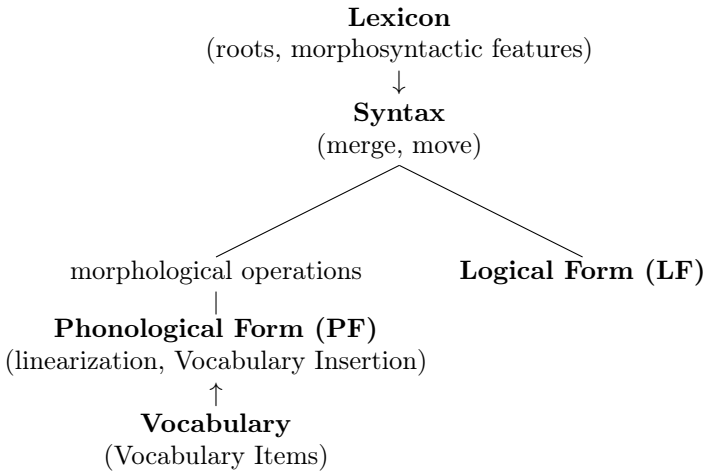
The research presented in this thesis is situated within the theoretical framework of *Minimalism* (Chomsky 1995 and subsequent work). Specifically, I assume a model of grammar that takes syntax to be a purely derivational system that builds hierarchical representations of lexical items by means of the operations (*external*) *merge* and *move* (i.e. *internal merge*). The operation *merge* takes two lexical elements out of the lexicon and combines them into a larger unit. The operation *move* takes a lexical element that is already present in the derivation and remerges it in a higher position. This entails that a moved element leaves a *copy* of itself, rather than a *trace*: the Copy Theory of Movement

[7] In this thesis, I abstract away as much as possible from phonological or lexical differences between Standard Dutch and dialectal or colloquial Dutch. Only when they are relevant for the discussion at hand I will indicate such differences.

(revived by Chomsky 1993; this theory will turn out to be of significant importance for the analysis of pronominal doubling in chapter 2). At a point called Spell-Out, the derivation branches off to the interface levels Phonological Form (PF) and Logical Form (LF), which provide instructions for the *articulatory-perceptual* (A-P) and the *conceptual-intentional* (C-I) systems respectively.

I assume that this T/Y-model of grammar (Chomsky 1995) is combined with a *late insertion* model of morphology (e.g. *Distributed Morphology*, cf. Halle and Marantz 1993, 1994, Harley and Noyer 1999), according to which phonological and morphological information becomes available only *after* the syntactic component finished the derivation. Specifically, syntax operates on roots and morphosyntactic feature bundles that are taken from the *Lexicon*. After syntax is finished manipulating these elements, but before the level of PF is reached, morphological operations (like *morphological reanalysis*, cf. chapter 2) may apply to the syntactic structure. Finally, at the level of PF, linearization of the hierarchical structure takes place, and the abstract feature bundles are replaced by Vocabulary Items (VIs). VIs are lexical items – with a phonological exponent and information about where the item can be inserted – that are stored in the *Vocabulary*. A representation of this model of grammar is given in (5).

(5) The model of grammar



This brief introduction of the model of grammar that I assume throughout this thesis suffices for now. More specific concepts, claims and assumptions will be introduced at the point where they become relevant to the discussion at hand.

1.2.2 Microvariation

The study of syntactic microvariation has gained a lot of ground in the last two decades (cf. Kayne 2000, 2005, Barbiers 2009, and the collection of papers

in Barbiers et al. 2008b).[8,9] An important reason for this is that by studying closely related languages or dialects it is possible to keep most variables constant while focussing on the variable under study. As Kayne (2000:5) puts it:

> *If it were possible to experiment on languages, a syntactician would construct an experiment of the following type: take a language, alter a single one of its observable syntactic properties, examine the result to see what, if any, other property has changed as a consequence. If some property has changed, conclude that it and the property that was altered are linked to one another by some abstract parameter. Although such experiments cannot be performed, I think that examining pairs (and larger sets) of ever more closely related languages, one can begin to approximate the results of such an experiment.*

Within the theoretical framework of *Minimalism* (cf. section 1.2.1), it is assumed that syntax is invariable (i.e. syntactic principles are constant across languages), and that all syntactic variation should be reduced to the lexicon (variation in morphosyntactic features) or the level of PF (variation in spell out/lexicalization). Although most of the syntactic microvariation discussed in this thesis can indeed be reduced to the lexicon or PF, I will show in chapter 2 that part of the variation regarding doubling needs to be accounted for in syntax (*pace* Chomsky 1995), namely in terms of optional subextraction or pied piping (cf. Koopman and Szabolcsi 2000, Barbiers et al. 2009).[10]

In the context of *microvariation* in Dutch, I distinguish *colloquial* Dutch from *dialectal* Dutch (see also section 1.3.2). With colloquial Dutch I refer to spoken language that does not have a clear geographic distribution but may occur in the whole Dutch speaking language area. Dialectal Dutch on the other hand, has a clear restricted regional distribution. Especially for phenomena that do not have a clear geographic distribution (colloquial Dutch) – like different patterns of pronominal doubling in long-distance A-bar dependencies (cf. chapter 2) – it is generally unclear *why* an informant (or variety) chooses one variant and not another.

Barbiers (2006, 2009) takes there to be a distinction between *ungrammatical* structures and *unrealized* structures. Ungrammatical structures are structures that violate (specific instances of) general syntactic principles. Unrealized structures, on the other hand, are structures that can be generated by the grammar,

[8] Cf. recent large scale dialect syntax projects like the SAND project (*Syntactic Atlas of the Dutch Dialects*, Barbiers et al. 2005, 2008a), and the Edisyn project (*European Dialect Syntax*, cf. `http://www.dialectsyntax.org`).

[9] See Biberauer (2008) for a comprehensive overview and discussion of different models of (the limits of) syntactic variation more generally – e.g. *functionalism* (predominantly communicative and/or processing considerations) vs. *formalism* (UG, Principles and Parameters) – with special focus on the parametric enterprise (within *Minimalism*).

[10] Of course, this claim immediately raises the question if it is possible to reduce (the effects of) subextraction/pied piping to a PF phenomenon, so that the minimalist assumption about the locus of microvariation can be maintained. I will discuss this at the end of chapter 2.

but that are simply not used in a certain variety. Under this perspective, the *why* question – namely why in a given variety a certain grammatical structure is realized or unrealized – is to be answered in terms of sociolinguistic circumstances, and not in terms of language internal properties. In this thesis, I will primarily focus on the difference between *grammatical* and *ungrammatical* structures, and I will remain largely agnostic about the sociolinguistic variables that may influence the *(non-)realization* of a given grammatical structure.

1.2.3 Doubling

This thesis restricts its attention to *syntactic* doubling phenomena. The definition of syntactic doubling that I assume throughout this thesis is given in (6). From the perspective of principles like *Compositionality* (attributed to Frege 1892) and *Economy* (Chomsky 1995 among many others) – according to which natural language should be maximally economical and each step in the syntactic derivation should contribute to the semantics – the existence of pure syntactic doubling (i.e. doubling that does not have a semantic effect) is highly unexpected. This immediately raises the question of *why* syntactic doubling exists in the first place. Because this thesis is mostly concerned with providing a principled account of doubling, and less so with offering a full fledged answer to the *why* question, only in passing it will offer some speculation on the *function* of doubling in natural language.[11]

(6) *syntactic doubling*: overt repetition of one or more semantically redundant morphosyntactic feature(s)
[adapted from Barbiers 2008, Alber 2008:142]

The definition in (6) makes sure that the repetition of a feature *bundle* as well as the repetition of a *single* feature is considered doubling – as long as this repetition does not have semantic import. This ultimately means that *agreement* – the expression of a single feature on two different elements, e.g. subject-verb agreement as in *de taalkundig**en***$_{[PL]}$ *lop**en***$_{[PL]}$ 'the linguists walk' – is a core case of doubling (notice this is true only when agreement is assumed not to have a semantic effect).

Both doubling of a feature bundle and doubling of a single feature figure prominently in this thesis, although I will not be concerned so much with the phenomenon of agreement in the traditional sense, like the subject-verb agreement mentioned before. Doubling of a full feature bundle is represented by identical doubling of a pronoun in long-distance A-bar dependencies. This is illustrated for a root *wh*-question in (7), in which the pronoun *wie* appears twice without making any contribution to the semantics of the clause. Assuming that

[11] Doubling, and variation more generally, might be caused by general principles that are not specific to the faculty of language ('the third factor of language design', cf. Chomsky 2005:6), e.g. principles of data analysis and principles of efficient computation. Alber (2008) for example argues that doubling in long-distance A-bar movement in Tyrolean facilitates processing.

pronouns have internal structure and contain several morphosyntactic features (like ϕ-features, cf. section 2.5.2), the doubling construction in (7) is a case of doubling of a feature bundle.

(7) **Wie** denk je **wie** het gedaan heeft?
who think you who it done has
'Who do you think has done it?' [colloquial Dutch]

Doubling of a single morphosyntactic feature is presented here by non-identical doubling involving a complex *wh*-phrase. This is illustrated for a root *wh*-question in (8), in which the complex *wh*-phrase remains in the left periphery of the lower clause whereas the pronoun *wat* 'what' surfaces in the left periphery of the higher clause. This sentence has exactly the same meaning as the Standard Dutch variant in (9). Assuming that the *wh*-feature on the *wh*-phrase *welke man* 'which man' is doubled by the pronoun *wat* 'what' in the higher clause in (8) (*wat* being completely underspecified for other features, cf. section 2.5.2.2), the doubling construction in (8) is a case of doubling of a single feature.

(8) **Wat** denk je **welke man** het gedaan heeft?
what think you which man it done has
'Which man do you think has done it?' [colloquial Dutch]

(9) **Welke man** denk je **dat** het gedaan heeft?
which man think you that it done has
'Which man do you think has done it?' [Standard Dutch]

In sum, both *full* and *partial* overlap in morphosyntactic features between two elements are considered cases of doubling.

The definition of *syntactic doubling* as given in (6) excludes *reduplication*, which is

> *a morphological process relating a base form of a morpheme or stem to a derived form that may be analyzed as being constructed from the base form via the affixation (or infixation) of phonemic material which is necessarily identical in whole or in part to the phonemic content of the base form.* [Marantz 1982:437]

Since reduplication generally contributes to the semantics of a construction (e.g. pluralization, intensification, iterativity) – as illustrated in (10) for nominal reduplication forming plurals – it does not fall under the definition of *syntactic doubling* in (6), and will therefore not be discussed any further.[12]

[12] As pointed out to me by Marcel den Dikken, reduplication might very well be a PF phenomenon: a certain feature that is present in syntax (e.g. [plural]) is linked to phonological reduplication in the PF component. If reduplication is a PF phenomenon (and thus not the spell out of the repetition of a feature in syntax, i.e. not *syntactic doubling*), reduplication itself does not contribute to the semantics, because PF does not feed into LF. It should then in principle be possible to find reduplication in contexts in which the feature that triggers reduplication is not related to an LF interpretation.

(10) a. ɟimurU ‘house’
ɟimuɟimurU ‘houses’

b. gindalba ‘lizard sp.’
gindalgindalba ‘lizards’

[Dixon 1977, as cited in Marantz 1982:453]

1.3 Empirical background

The data that form the backbone of this study come from the SAND corpus (*Syntactische Atlas van de Nederlandse Dialecten* ‘Syntactic Atlas of the Dutch Dialects’, Barbiers et al. 2005 (SAND1), Barbiers et al. 2008a (SAND2))[13] and two large scale questionnaire studies: the *Meertens Panel questionnaires* (MPQ1 and MPQ2). This section briefly presents the methodology and design underlying these data collections.

1.3.1 SAND data

The SAND corpus is the result of a large scale collaboration (starting in 2000) between linguists of the Meertens Instituut (Amsterdam), the Fryske Akademy (Leeuwarden), the Universities of Amsterdam (UvA) and Leiden (the Netherlands), and the Universities of Gent and Antwerpen (Belgium). The SAND data give a detailed overview of the (geographic distribution of) *syntactic* variation in several linguistic phenomena in 267 locations in a part the Dutch speaking language area: the Netherlands, the Dutch speaking part of Belgium (Flanders) and a small part of northwest France.[14]

1.3.1.1 Methodology

The research and fieldwork conducted in the SAND project consisted of four different stages. In the first stage, an inventory of syntactic variation was made, mainly on the basis of literature research. In the second stage, a *written* questionnaire was sent to 368 informants (from the database of the Meertens Instituut), with the goal to get a first insight into the geographic distribution of the different syntactic phenomena. On the basis of the results of the written questionnaire study, a motivated and efficient choice was made for the kind of test sentences and the locations in which these test sentences were to be questioned during the oral interviews. These oral interviews comprised the third stage of research, and were carried out in 267 locations in the Dutch speaking language area (cf. map 1.1). In the oral interview round, informants were presented with

[13] See the DynaSAND (Barbiers et al. 2006) for the complete SAND database (`http://www.meertens.knaw.nl/sand/`). All the maps in this thesis are generated with the free mapping software that is available on the DynaSAND webpage.

[14] Dutch is also (one of) the official language(s) of Curaçao, Bonaire, Sint Maarten, Sint Eustatius, Saba (the former Netherlands Antilles), Aruba and Surinam – all former colonies of the Netherlands.

a test sentence and asked to indicate whether such a sentence occurs in their dialect (*indirect grammaticality judgment task*) and how it should be translated into their dialect (*translation task*). To ensure that informants were not influenced by (the lexical and phonological properties of) the standard language spoken by the interviewer, there were always at least two informants present when the interview took place. This design led to two scenarios. In the first, the two informants could discuss the sentences between them (Belgium); in the other scenario, one of the informants was trained to be the interviewer (the Netherlands). The fourth and final stage of research consisted of telephonic interviews (with only part of the informants) in order to complete missing or unreliable data from the oral interview round. See Cornips and Jongenburger (2001a,b), and Barbiers and Bennis (2007) for more details on the methodology of the SAND project (cf. also Cornips and Poletto 2005 on syntactic elicitation techniques more generally).

For the atlases, only the data obtained through the oral and (when available) telephonic interview rounds are used, as those data are the most reliable (cf. Cornips and Jongenburger 2001a,b). Similarly, in the presentation and discussion of the SAND data on relativization in section 4.5, I will make use only of the oral data collection. It should further be noted that, since I reanalyzed the original data as published in SAND1 (Barbiers et al. 2005), the maps presented in section 4.5 sometimes differ (albeit slightly) from the ones that are published.

1.3.1.2 Speakers and locations

The choice for the (number of) measuring points (locations) was based on the desire to have an evenly spread of the measuring points over the language area, the expected amount of variation in a given area, and the history of a given location/area. On the basis of these criteria, a total number of 267 measuring points was chosen: 158 locations in the Netherlands, 102 locations in Belgium, and 7 locations in France, as can be seen on the following map.

Figure 1.1: SAND measuring points

To keep factors that may possibly contribute to social variation as constant as possible, the informants themselves were selected on the basis of the following six (social) criteria: (i) the informant speaks the dialect of the location in which he/she currently lives, (ii) the informant as well as his/her parents are borne and raised in the location in which the dialect is spoken, (iii) the informant has not lived elsewhere for longer than 7 years (and only after the age of 18), (iv) the informant speaks dialect at home and in at least one other public domain, (v) the informant is between 55 and 70 years old, and (vi) the informant preferably belongs to the lower class or lower middle class. See Cornips and Jongenburger (2001a,b) for more details on the selection and recruitment of informants.

1.3.2 Meertens Panel questionnaire data

In addition to making use of the SAND data, I carried out two large scale online questionnaire studies to complete the existing data set, henceforth the Meertens Panel questionnaire (MPQ) studies. Most of the constructions that were tested in the MPQs, were variants of constructions that were not tested in the SAND project, or constructions that came up as spontaneous translations of test sentences in the SAND project but that were not explicitly tested with all SAND informants. Whereas the test sentences in the MPQs were thus based on a thorough evaluation of the SAND test sentences, the MPQs are not a continuation of the SAND project in the sense that geographic distribution was

not as relevant a factor as it was in the SAND project. More specifically, most constructions that I tested – especially doubling in long A-bar dependencies – are known not to show a clear geographic distribution (colloquial Dutch). The dialect background of the informants – most prominently the location – was thus less relevant. Therefore, I chose to make use of the large group of respondents obtained by the Meertens Instituut: the *Meertens Panel*.[15] The Meertens Panel consists of people who are interested in Dutch language and culture and signed up electronically for the Meertens Panel to answer digital surveys. The only requirements candidates have to meet are that they be at least sixteen years old and are currently living in the Dutch speaking language area.[16]

Because in the MPQs the test sentences are presented to the informants in *spoken form* (cf. *infra*), these questionnaire studies represent a methodologically rather new and advanced way of eliciting (large numbers of) introspective grammaticality judgments.

1.3.2.1 Design of the questionnaire

The first Meertens Panel Questionnaire (MPQ1) was carried out in November-December 2010, and the follow up questionnaire (MPQ2) was carried out more than a year later, in January 2012. Both questionnaires had exactly the same design to make cross comparison of data possible, but the informants were different (although there may have been partial overlap).

The questionnaires were offered to the informants online, in spoken form. That is, informants were confronted with spoken sentences without a written version of the sentence on the computer screen. This way, informants were not exposed to the more formal written language, thereby reducing the chance of them giving *normative judgments*.[17] The test sentences were all recorded by the same Dutch speaker and pronounced with the most natural intonation, to simulate spoken Dutch as well as possible. The instruction to the questionnaire pointed out that in case a certain word in a given test sentence is different from the word that an informant would use in his/her spoken Dutch, the informant should replace the word with the word used in his/her spoken Dutch, repeat the sentence to him/herself, and then judge the sentence. In addition, because of the length of the questionnaire, the introduction recommended informants to take a break once in a while.

The test sentences were presented to the informants in randomized order, not in blocks of two or more sentences, as a result of which the informants were unable to make a direct comparison between sentences (no direct *relative*

[15] Part of the Meertens Panel informants were also informants in the SAND project.

[16] See `http://www.meertens.knaw.nl/meertenspanel/` for more details.

[17] A recurrent problem in microcomparative research is that of *normativity*: it is often unclear whether informants provide true grammaticality judgments or whether their judgments are *normative judgments*, highly influenced by the normative standard language (cf. Cornips and Jongenburger 2001a,b, Barbiers 2009:1608 a.o.).

jugdments). The reason for this was twofold. First, I am not primarily concerned with relative grammaticality judgments between sentences, but more with whether or not a sentence occurs. Second, the task of comparing two or more test sentences that are only offered in spoken form (no written version on screen) is rather complicated. Informants were asked to indicate whether or not the given sentence occurs in their spoken Dutch (*ja* 'yes' or *nee* 'no'), and if so, how *gebruikelijk* 'common' it is on a scale from 1 to 5 (where 1 means *ongebruikelijk* 'uncommon' and 5 means *heel gebruikelijk* 'very common'). This indirect way of asking grammaticality judgments (i.e. informants are not directly asked whether or not a given test sentence is grammatical or acceptable) prevents the informants from answering too much from a normative perspective (cf. Cornips and Jongenburger 2001a,b).

As for the technical implementation of the questionnaire, answers were immediately saved so informants were able to take a break. It was made sure that informants could not proceed to the next sentence without having listened to the sound file and without having given a judgment to the current sentence, but informants were allowed to go back and possibly change their judgments. Not completely filled out questionnaires were excluded from analysis.

MPQ1 consisted of two different questionnaires (randomly distributed over the informants): questionnaire A about the left periphery of embedded questions and relative clauses, and questionnaire B about (doubling in) long-distance A-bar dependencies. MPQ2 consisted of two different questionnaires as well (randomly distributed over the informants): questionnaire A about doubling and intervention effects in long-distance A-bar dependencies, and questionnaire B about doubling in long-distance A-bar dependencies with multiple embeddings and doubling with prepositional phrases in long-distance A-bar dependencies. The total number of informants that responded to the questionnaires was rather high, but unfortunately so was the number of informants that could not be used for analysis. All informants who were not raised in the Dutch speaking language area, who did not indicate where they were raised or who responded incorrectly to the ungrammatical filler items (cf. *infra*), were excluded from analysis. This leaves the numbers as indicated in table 1.1.

Table 1.1: Facts MPQ1 and MPQ2

	MPQ1		MPQ2	
	A	B	A	B
# of test sentences	102	92	75	81
# of usable respondents	452	333	255	380

All questionnaires contained at least one ungrammatical filler item (in most cases two), namely an embedded *wh*-question that is introduced by the declarative complementizer, as illustrated in (11). Such sentences are known to be

truly ungrammatical.[18]

(11) a. * Ik vraag me af dat hij het gedaan heeft.
I wonder that he it done has

b. * Ze vroeg dat jij denkt wie het gedaan heeft.
she asked that you think who it done has

Between 3% and 13% of the informants in all questionnaires did the filler test(s) incorrectly, i.e. they accepted the ungrammatical test sentences. As mentioned before, all informants that responded incorrectly to one (or more) ungrammatical filler items were excluded from analysis. Because there were not many filler items in MPQ1 and MPQ2, the task of deciding speaker reliability is complicated. I take the fact that 3%-13% of the informants did the filler test(s) incorrectly to indicate that the status of sentences that are accepted by less than 15% of the informants *and* that do not show a clear geographic distribution (to make sure that we are not dealing with a dialectal/regional Dutch phenomenon) is marginal at best.[19] In this thesis I consider a test sentence that is accepted by less than 15% of the informants to represent an *ungrammatical* structure.

I am mostly concerned with whether or not a given sentence occurs in the Dutch speaking language area (*yes* or *no*), and less with how *common* such a sentence is judged to be (on a scale from 1 to 5) – the latter result is mostly interesting in the context of comparing the grammaticality judgments on different sentences by a single informant. Put differently, as I am mostly interested in phenomena that occur in colloquial Dutch, I look at the status of sentences not at the level of an individual informant, but at the level of the group of informants as a whole. Therefore, in the remainder of this thesis, I will use the following symbols to indicate what the status of a given sentence is in the whole Dutch speaking language area (unless indicated otherwise, e.g. in case of relative grammaticality judgments in a particular language variety).[20]

[18] Only MPQ1-A contained another type of ungrammatical filler item – in addition to an ungrammatical filler sentence of the type in (11) – namely an embedded *wh*-question introduced by the string *wie dat of* 'who that whether'. Such doubly filled Comp patterns are known to be non-existent.

(i) * Ik vraag me af wie dat of het gedaan heeft
I wonder who that whether it done has

[19] This relatively high number (15%) is further justified by the fact that the questionnaires were long, the test sentences were complicated, and the method of questioning was complex. These factors may all have contributed to noise in the data.

[20] The exact scores of the informants regarding the sentences in the MPQs (*no* (0) or *yes* (1-5)) are thus not directly converted into grammaticality judgments like ?, ??, ?* etc.

(12) a. % = colloquial Dutch (i.e. accepted by at least 15% of the speakers and no clear geographic distribution)
b. = accepted by all informants (Standard Dutch)
c. * = not attested
d. ?* = very marginally attested (considered ungrammatical)
e. ?? = unclear status (for whatever reason)

Finally, let me point out that the constructions that were tested in the MPQs show a high degree of variability. Part of this variability is systematic, but it might very well be the case that part of this variability is the result of *methodological noise* as well. At least part of this methodological noise could potentially be avoided by (additional) large-scale *corpus research*. However, because most of the constructions that were tested in the MPQs are very complex – in particular doubling in long-distance A-bar dependencies – they are unsuitable for corpus research, as their frequency in spoken language is too low (whereas their variability is high) to be adequately researched. Notice furthermore that although corpus-based research could in principle *complement* the MPQ results, it could never *replace* questionnaire-based research: with corpus-based research it is impossible to establish the *limits of variation*. Put differently, corpora only provide information about the occurrence of sentences, not about their non-occurrence.

Some of the methodological problems that a questionnaire study faces could also be overcome by face-to-face *oral interviews*, i.e. structured in-depth oral interviews guided by a questionnaire. More specifically, oral interviews, as opposed to questionnaire studies, allow better control over the quality of the elicited data (cf. Cornips and Jongenburger 2001b). First, it offers the interviewer the opportunity to observe the language of the informant, and to make sure that the informant understands the task it is faced with. If, for example, a sentence is unclear to the informant (for whatever reason), it is possible to provide additional information on the fly, like context. Second, the interviewer can ask the informant further questions about a given construction, for example in case it is unclear if the answer of the informant is reliable. Because carrying out oral interviews is very time consuming, it was not feasible to do within the scope of this thesis. In order to get the most reliable data possible and to evaluate the reliability of the existing data, the questionnaire-based research presented in this thesis should be complemented with oral interviews in the future.

1.3.2.2 Speakers and locations

The Meertens Panel informants are not as orderly distributed over the Dutch speaking language area as the SAND informants are. That is, the geographic spread is almost entirely restricted to the Netherlands (Belgium is severely underrepresented), with the western-central part of the Netherlands best represented. Maps 1.2 and 1.3 show the geographic spread of the informants in

MPQ1 and the informants in MPQ2 respectively. More specifically, these maps indicate the locations where the informants were raised (not the location of their current residence). The circle indicates these locations for the informants that participated in the A-questionnaire, and the plus sign indicates these locations for the informants that participated in the B-questionnaire. Notice that each symbol on these maps represents a location for which there are data, but it does not indicate the *density* of informants (i.e. the number of informants per location) in the sample. It is therefore important to keep in mind that a map that is based on the MPQ data and that depicts more than one linguistic variant, always shows variation per *location*, not necessarily variation within an *individual* speaker (although individual speaker variation may be there as well). Table 1.2 gives the number of speakers per location for all locations that have 5 or more speakers in the sample.

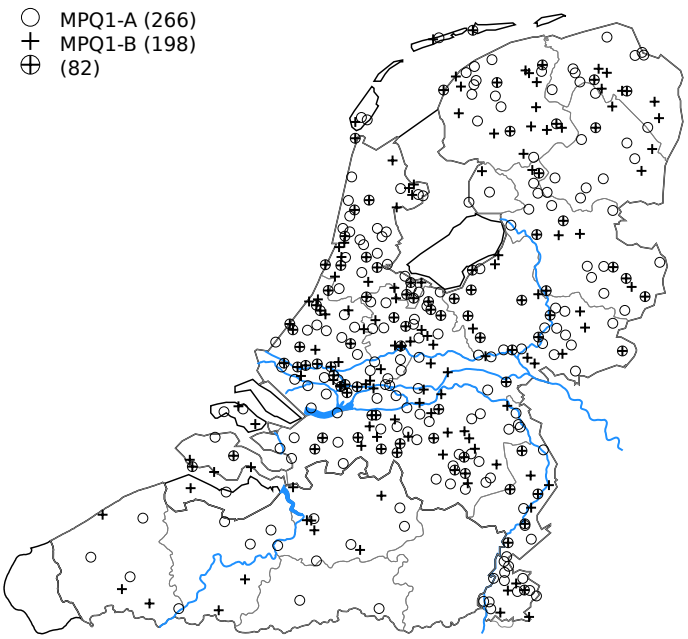


Figure 1.2: Geographic spread of informants MPQ1

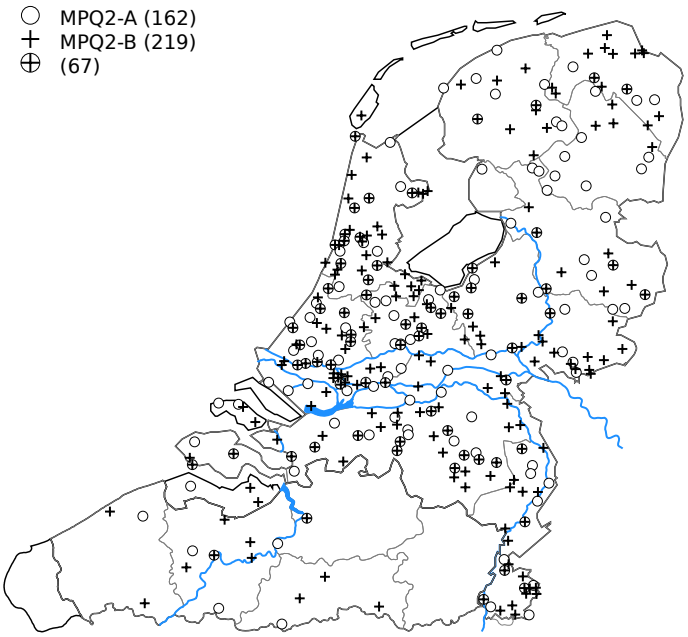


Figure 1.3: Geographic spread of informants MPQ2

Table 1.2: Locations in MPQs with at least 5 informants

	Amsterdam	Den Haag	Rotterdam	Utrecht	Haarlem	Breda	Groningen	Delft	Tilburg	Leeuwarden	Maastricht	Nijmegen	Eindhoven	Dordrecht	Apeldoorn
1-A	39	23	8	10	7	4	4	5	4	4	8	4	4	5	2
1-B	19	14	14	6	7	7	7	5	5	5	0	5	4	3	5
2-A	28	14	6	2	7	9	2	3	3	0	2	3	1	0	3
2-B	35	12	14	7	9	5	4	4	4	7	5	3	5	5	3

The profile of the informants in the MPQs was rather different from the profile of the informants in the SAND project. Whereas the informants in the SAND project were carefully selected on the basis of several criteria, this was less relevant (cf. *supra*) and not possible for the informants in the MPQs, as there were not always socio-biographical data available. The only criterion that I used in selecting the MPQ informants was that they were raised in the Dutch speaking language area. A consequence of this methodological choice is that the MPQ data (unlike the SAND data) really reflect the spoken Dutch language *in all its variability*

1.4 Outline and outlook

Chapter 2: Doubling and the structure of A-bar chains

In chapter 2, I propose a unified account of doubling in long-distance A-bar dependencies, on the basis of new empirical data on pronominal doubling in long-distance relative clauses (RCs) and pronominal doubling in long-distance (embedded) *wh*-questions (*wh*-Qs). I argue that all long-distance A-bar dependencies are derived by successive-cyclic movement via SpecCP of (part of) the A-bar pronoun (subextraction vs. pied piping), and that doubling comes about by spelling out more than one (part of a) copy in the A-bar chain at PF. The specific patterns of pronominal doubling in *wh*-Qs and in RCs (and the differences between them) are the result of the different unique feature specifications, and thus different lexicalization options, of the A-bar pronouns involved (*d*-pronouns and *w*-pronouns).

The empirical contribution of this chapter – when compared to existing literature and available data on pronominal doubling – is a thorough description of the (limits of) variation in doubling patterns in long-distance A-bar dependencies in the Dutch speaking language area, with special focus on RCs. The availability of the large set of new empirical data from the MPQs makes it possible to review, reaffirm or sometimes even prove false existing empirical claims, and, furthermore, to generate new empirical claims. The fundamental theoretical contribution of this chapter is the development of a systematic unified analysis of two phenomena that so far have been treated independently from each other: pronominal doubling in long-distance RCs and pronominal doubling in long-distance *wh*-Qs.

Chapter 3: Doubling and the syntax of relative clauses

In the first part of chapter 3, I show that doubling in colloquial Dutch long-distance RCs poses a challenge for the currently most prominent analyses of RCs: Head Internal Analyses (HIAs, *raising* or *matching*). According to HIAs the RC head originates in the gap position inside the RC and moves to the left periphery. I propose a specific implementation of the traditional Head External Analysis (HEA), according to which the RC head is base-generated in the highest SpecCP position of the RC itself (but crucially not in the gap position inside the RC, as in HIAs). Besides adequately accounting for doubling in long-distance RCs, this analysis also accounts for all attested variation in the left periphery of Dutch RCs, notably doubly filled Comp patterns.

The second part of chapter 3 makes a case for a HEA of RCs. It shows that although HIAs have gained more and more ground in recent years, a HEA fares better in many respects. Furthermore, in line with existing literature, I demonstrate that the argument that has always been taken to strongly argue against a HEA, namely connectivity effects between (material inside) the RC head and the RC internal gap, is not very well founded. Connectivity effects

are not always a proper diagnostic for movement (of the RC head), as a result of which their presence or absence in RCs cannot be used as an argument in favor or against either a HEA or a HIA of RCs.

Empirically, the contribution of this chapter lies in the observation that a new set of empirical data, namely pronominal doubling in long-distance RCs, provides a new perspective on an old issue within generative linguistics: the syntax of relative clauses. Related to this observation is the main theoretical contribution of this chapter: the reinstatement of an old idea, namely the idea that the RC head does *not* originate in the gap position inside the RC.

Chapter 4: On relative pronouns and complementizers

Chapter 4 explores the nature of relative pronouns, the other functions that such relative pronouns may have (multipurpose pronouns), and the relation between relative pronouns and complementizers. I reject a homophony analysis to the different functions of a multipurpose pronoun and argue instead that a multipurpose pronoun has a single underspecified lexical entry, as a result of which it may appear in more than one syntactic environment and part of its meaning is determined contextually or configurationally (null hypothesis). In the first part of this chapter, I set the basis for a new implementation of this underspecification theory to multipurpose pronouns. The second part of the chapter explores the relation between (relative) pronouns and complementizers. It argues that even though the Dutch finite declarative complementizer and the neuter gender distal demonstrative/relative pronoun are not the same lexical item, they are diachronically related in a way that is compatible with the null hypothesis. It is the most underspecified *d*-pronoun that served as a source for grammaticalization into a complementizer. In addition, I argue against the recently popular claim that complementizers are in fact relative pronouns. This claim is further challenged by a case study on the distribution of pronouns and complementizers in long-distance RCs in some southern Dutch varieties, with special focus on subject/object asymmetries.

Chapter 5: Conclusions

The final chapter offers a summary and highlights the most important empirical and theoretical conclusions and contributions of this thesis.

CHAPTER 2

Doubling and the structure of A-bar chains

2.1 Introduction

Over the past few decades, doubling in long-distance *wh*-questions as illustrated in (13) for colloquial Dutch – traditionally referred to as *wh*-copying (13a) and *wh*-scope marking (13b) – has received considerable attention (cf. the volume by Lutz, Müller, and von Stechow 2000 a.o.).

(13) a. % **Wie** denk je **wie** het gedaan heeft?
who think you who it done has
'Who do you think has done it?'

b. % **Wat** denk je **wie** het gedaan heeft?
what think you who it done has
'Who do you think has done it?'

There is not much consensus on the status of (the chain(s) in) such long-distance A-bar dependencies. Whereas *identical* doubling as in (13a) is often treated as multiple spell out of chain links (cf. Fanselow and Mahajan 2000, Nunes 2004, Felser 2004, Bruening 2004, 2006, Bošković and Nunes 2007, Barbiers et al. 2009 a.o.), there is an ongoing debate about *non-identical* doubling as in (13b) (cf. Felser 2001, Fanselow 2006 for an overview of different analyses of *wh*-scope marking).

Situated against this background, this chapter presents a novel account of doubling in long-distance A-bar dependencies, on the basis of new empirical data on pronominal doubling in colloquial Dutch long-distance *relative clauses*

(RCs).[1] An example of (identical) doubling in a RC is given in (14).

(14) % Dat is de man [$_{RC}$ **die** ik denk [**die** het gedaan heeft]].
that is the man RP I think RP it done has
'That is the man who I think has done it.'

Since Chomsky (1977), it is standardly assumed that the syntax of *wh*-questions (*wh*-Qs) and the syntax of RCs are related, in the sense that both involve A-bar (or *wh*-)movement.[2] I therefore start from the presupposition that, despite superficial differences, doubling in RCs as well as doubling in *wh*-Qs are all of a piece, and I present a *unified analysis* of these phenomena, that so far have been treated independently from each other (cf. Boef 2008a, 2012b, Koopman and Sportiche 2008 for doubling in RCs, and cf. Barbiers et al. 2009, den Dikken 2009b a.o. for doubling in *wh*-Qs).

I take all long A-bar dependencies to be derived by successive-cyclic movement via SpecCP of the interrogative/relative pronoun (henceforth *A-bar pronouns*, following van Kampen 1997 and later work). For linearization purposes, all copies but the head of the movement chain of the pronoun must delete at PF, in accordance with the LCA (Kayne 1994). This is illustrated in (15), where strikethrough indicates PF deletion/non-realization.

(15) [$_{CP}$ **pronoun**$_1$ ... [$_{CP}$ ~~pronoun$_1$~~ ... ~~pronoun$_1$~~ ...]]
no doubling

In certain cases the intermediate copy of the pronoun may escape this linearization requirement, as a consequence of which it can be spelled out, in addition to the head of the chain (Nunes 2004). This results in *identical doubling*, as illustrated in (16) and exemplified by (13a) for a *wh*-Q and by (14) for a RC.

(16) [$_{CP}$ **pronoun**$_1$ ... [$_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
multiple copy spell out: identical doubling

I propose that the internal structure of A-bar pronouns includes an *operator* – located in the specifier of the pronoun (cf. Szabolcsi 1994) – which becomes PF visible when extracted. When an A-bar pronoun in a long-distance A-bar dependency has reached the embedded CP domain, the following two possibilities emerge. Either the whole pronoun (containing the operator that triggers movement) moves up (resulting in (15) or (16)), or only the operator itself moves up – the pronoun and the operator in its specifier being *equally local* to the higher SpecCP. This latter scenario will be referred to as *subextraction* of the operator (i.e. the lack of *pied piping* of the full pronoun). When the operator subextracts,

[1] Earlier versions of parts of this chapter were presented at the OC colloquium (Radboud University Nijmegen, November 2011) and at ConSOLE XX (University of Leipzig, January 2012).

[2] The claim that A-bar movement in *wh*-Qs may have a different landing site than A-bar movement in RCs (cf. Rizzi 1997 a.o.) is irrelevant to the subject of this chapter, but see chapter 3 for some discussion on this matter.

it is spelled out in its final landing site, the higher SpecCP. Since deletion of the pronoun that is left behind by subextraction of the operator in the lower SpecCP would lead to a *recoverability* problem, it needs to be spelled out. In doing so, a violation of the *Condition on Extraction Domain* (CED, Huang 1982) or the *Freezing Principle* (Wexler and Culicover 1980) is circumvented. This particular means to salvage an otherwise illicit step in the derivation (cf. van Craenenbroeck and van Koppen 2008), I will refer to as *rescue by PF spell out* – the logical counterpart of *rescue by PF deletion* (Bošković 2011). As spell out of the pronoun *subsumes* spell out of the operator (i.e. A-bar pronouns spell out phrases, cf. Weerman and Evers-Vermeul 2002, Barbiers et al. 2009 a.o.), the intermediate chain link will always surface as a full pronoun, as illustrated in (17).

(17) $[_{CP}$ **operator**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
subextraction plus double spell out: non-identical doubling

The operator gets spelled out as *wat* 'what' by default, as exemplified by (13a) – *wat* being the most underspecified A-bar pronoun in Dutch (cf. Postma 1994, Bennis 1995 a.o.).

My analysis of doubling in long A-bar dependencies is highly inspired by the analysis of Barbiers, Koeneman, and Lekakou (2009) (henceforth BKL) regarding doubling in Dutch long-distance root *wh*-Qs. However, whereas many aspects of BKL's analysis of doubling also feature prominently in my analysis of doubling, my analysis overcomes some of the problems their analysis faces. Furthermore, whereas BKL's analysis does not carry over to doubling in RCs, my analysis provides a unified account of the doubling patterns in RCs and *wh*-Qs. The crucial innovation of my analysis is that the variation and differences in the doubling patterns in RCs and *wh*-Qs are accounted for by the feature specifications of the A-bar pronouns involved (namely *wat* 'what', *wie* 'who', *die* 'that.C' and *dat* 'that.N'), and different lexicalization options and requirements. More specifically, in many cases more than one A-bar pronoun may lexicalize the (copy of) the interrogative/relative pronoun in *wh*-Qs and RCs, resulting in variation (recall that I assume a late insertion model of morphology, cf. section 1.2.1). In RCs, variation is furthermore caused by the choice of which features of the relative pronoun are spelled out: *syntactic* gender (common/neuter) or *semantic* animacy (roughly human/non-human). Whereas Standard Dutch requires the spell out of syntactic gender, in colloquial Dutch semantic animacy may be spelled out instead. Finally, the choice of which pronoun is inserted is dependent on the nature of the clause: *wh*-Q or RC, i.e. there is a *wh*-requirement on the introduction of *wh*-Qs, as a result of which *d*-pronouns cannot introduce *wh*-Qs.

Doubling in RCs and doubling in *wh*-Qs thus receive a unified account in this chapter. Variation in doubling in the Dutch speaking language area can be reduced to the availability of multiple copy spell out (doubling vs. no doubling), the availability of subextraction (non-identical doubling vs. identical doubling

or no doubling), and the choice for the lexicalization of the A-bar pronoun (specific patterns of doubling).

This chapter is organized as follows. As my analysis of doubling in long-distance A-bar dependencies builds and improves on the analysis of doubling in long-distance root *wh*-Qs as proposed by BKL, section 2.2 first presents the relevant doubling data in *wh*-Qs and discusses their analysis of these data, along with a number of other theoretical notions and assumptions that are instrumental to the analysis to be proposed. Section 2.3 introduces the doubling patterns in long-distance RCs and shows why BKL's analysis does not extend to cover those doubling data. In section 2.5, I outline my proposal for a unified account of doubling in long A-bar dependencies in detail. Section 2.6 provides (empirical) support for the proposed analysis, and in section 2.7, I extend the application domain of the analysis to cover other types of doubling as well: doubling involving complex *wh*-phrases and doubling of prepositional phrases. Section 2.8 sums up and concludes the chapter. The appendices at the end of this chapter discuss alternative analyses that take a *multiple chain* or *indirect dependency* approach to doubling in long A-bar dependencies (appendix A) and some remaining doubling data (appendix B).

2.2 Doubling in long-distance *wh*-questions

This section presents the attested pronominal doubling patterns in long-distance *wh*-Qs, and discusses the analysis of these doubling patterns as proposed by BKL. I first present the doubling patterns in section 2.2.1. Then I briefly discuss a proposal by Nunes (2004) regarding the conditions and requirements on the spell out of chains in section 2.2.3, as the analysis of BKL – and ultimately my analysis of doubling as well – relies in part on this proposal. Finally, section 2.2.4 presents and discusses the analysis of BKL in detail.

2.2.1 The data

In Standard Dutch long-distance (embedded) *wh*-Qs, the higher clause is introduced by the interrogative pronoun and the lower clause is introduced by the finite declarative complementizer *dat* 'that', as exemplified in (18). In all varieties of Dutch, the left periphery of finite embedded clauses needs to be introduced by at least one overt element. In the lower clause of a long-distance A-bar dependency in Standard Dutch this is always the declarative complementizer, i.e. Dutch does not show a *that*-t effect (e.g. Perlmutter 1971, Chomsky and Lasnik 1977) in terms of the presence or absence of the complementizer: the complementizer is always obligatorily present, independently of subject or object extraction. See section 2.3 for some more discussion on this matter. The construction in (18) is accepted by all Dutch speakers.[3]

[3]I abstract away from speakers for whom doubly filled Comp is obligatory, e.g. *Ze vroeg wie *(dat) jij denkt dat het gedaan heeft* 'she asked who *(that) you think that has done it',

(18) **Wie** denk je **dat** het gedaan heeft?
who think you that it done has
'Who do you think has done it?'

In this section, I will only be concerned with *wh*-Qs that question a person (*wie* 'who'). Furthermore, as the MPQ1-B data show that the (doubling) patterns in long-distance *embedded wh*-Qs – e.g. *Ze vroeg wie jij denkt dat het gedaan heeft* 'She asked who you think has done it' – are identical to the (doubling) patterns in *root wh*-Qs (as attested in the SAND corpus, Barbiers et al. 2005), I will not distinguish between the two constructions in this chapter. For ease of exposition, all doubling will be presented in root *wh*-Qs.

The sentences in (19) show all variants of the Standard Dutch sentence in (18) that are attested in the Dutch speaking language area.[4,5] These doubling configurations are reported to be semantically equivalent to their non-doubling counterparts: pronoun doubling in long-distance *wh*-Qs elicits a single answer response (these constructions are not multiple *wh*-Qs). I consider the constructions in (19) to be colloquial Dutch, because the MPQ1-B data do not show a clear geographic distribution for these constructions. This is illustrated on map 2.1. The numbers in between the first parentheses after the given doubling pattern indicate the number of *attestations* of that doubling pattern out of a total of 333 attestations, and the number in between the second parentheses indicates the number of *locations* in which the given doubling pattern is attested.[6] It is important to point out that the attestation of a certain combination of doubling patterns in a given location does not always indicate that a single informant allows that particular combination of doubling patterns, as there is no one-to-one relation between a location and an informant (i.e. there can be more than one informant per location, cf. section 1.3.2 for details).

(19) a. % **Wie** denk je **wie** het gedaan heeft?
who think you who it done has

b. % **Wie** denk je **die** het gedaan heeft?
who think you RP it done has

c. % **Wat** denk je **wie** het gedaan heeft?
what think you who it done has

d. % **Wat** denk je **die** het gedaan heeft?
what think you RP it done has
'Who do you think has done it.'

cf. SAND1 data (Barbiers et al. 2005:16), Boef (to appear).

[4]I follow BKL and gloss the element *die* as *relative pronoun* (RP), but this in no way means that I take the embedded clause to be a RC, nor does it mean that I take *die* to only be able to function as a relative pronoun. See chapter 4 for the nature of relative pronouns in general and the nature of pronoun *die* in particular.

[5]The pattern in (19c) (and (19d)) is better known as *wh-scope marking* or *partial wh-movement*. See Lutz et al. (2000), Felser (2001) and Fanselow (2006) amongst others for an overview of different analyses of *wh*-scope marking (cf. also appendix A to this chapter).

[6]As these numbers show, pattern *wat-die* is attested considerably less frequently than the other doubling patterns; the same is shown by the SAND1 data (Barbiers et al. 2005).

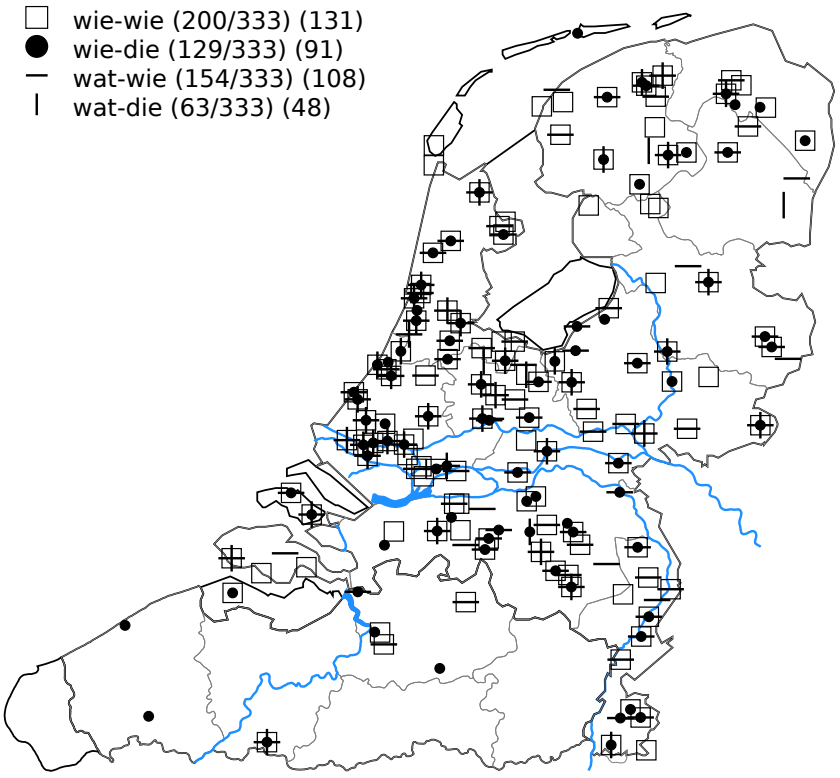


Figure 2.1: Doubling in embedded subject *wh*-Qs (MPQ1-B data)

The sentences in (20) show that if we were to switch the pronouns in the grammatical doubling patterns in (19), we get unattested outcomes.[7] This shows that there are clear limits on variation in doubling constructions.

(20) a. ?* **Wie** denk je **wat** het gedaan heeft?
who think you what it done has

b. * **Die** denk je **wie** het gedaan heeft?
RP think you who it done has

c. * **Die** denk je **wat** het gedaan heeft?
RP think you what it done has

[7]The doubling pattern in (20a) is attested very marginally in embedded *wh*-Qs in the MPQ1-B data (27/333=8%, but not in the SAND data), hence the ?* grammaticality judgment (cf. chapter 1).

2.2.2 Long-distance extraction

In the literature, there is not much consensus on the status of (the chain(s) in) long-distance A-bar dependencies in general, and long-distance *wh*-Qs in particular. Whereas *identical* doubling like (21a) is often treated as multiple spell out of chain links (see Fanselow and Mahajan 2000, Nunes 2004, Felser 2004, Bruening 2004, 2006, Bošković and Nunes 2007 amongst others), there is an ongoing debate about *non-identical* doubling, in particular the construction in (21b).

(21) a. % Ze vroeg **wie** jij denkt **wie** het gedaan heeft.
she asked who you think who it done has
'She asked who you think has done it.'

b. % Ze vroeg **wat** jij denkt **wie** het gedaan heeft.
she asked what you think who it done has
'She asked who you think has done it.'

The central debate revolves around the question of whether we are dealing with a *single* chain or with *multiple* chains (see appendix A to this chapter for some discussion on multiple chain analyses to (non-)identical doubling). A crucial issue within this debate is the status of *intervention effects*. For German, for example, it has been claimed that *wh*-scope marking (*was-wen*) patterns differently from long-distance *wh*-extraction (*wen-dass*) with respect to (negative) islands (cf. Dayal 1994, Felser 2001, 2004). This is illustrated in (22): only the *wh*-scope marking construction (22b) exhibits a negative island effect.

(22) a. **Wen** glaubst du *nicht* dass Maria getroffen hat?
who think you not that Maria met has
'Who don't you think Marie has met?'

b. * **Was** glaubst du *nicht* **wen** Maria getroffen hat?
what think you not who Maria met has

[German, Felser 2001:12]

As mentioned by BKL, it is sometimes overlooked that identical doubling patterns exactly like *wh*-scope marking, as illustrated for German in (23) (cf. Felser 2004, Rett 2006 a.o.). Put differently, there does not seem to be any difference between non-identical doubling (*wh*-scope marking) and identical doubling constructions in long-distance root *wh*-Qs in German with respect to intervening negation. Exactly the same holds for Dutch, as illustrated in (24).

(23) * **Wen** glaubst du *nicht* **wen** Maria getroffen hat?
who think you not who Maria met has

(24) a. **Wie** denk je *niet* dat zij uitgenodigd heeft?
who think you not that she invited has
'Who don't you think that she invited?'

b. * **Wat** denk je *niet* **wie** zij uitgenodigd heeft?
what think you not who she invited has

c. * **Wie** denk je *niet* **wie** zij uitgenodigd heeft?
who think you not who she invited has [BKL 2009:40]

Whatever the analysis of these intervention effects, the observation that non-identical doubling patterns exactly like identical doubling seems to suggest that the two doubling constructions should be treated in the same way. Leaving aside for the moment intervention effects (but see section 2.6.4, and see BKL, Szabolcsi and Zwarts 1993 for some discussion on the status and analyses of intervention effects), I assume that *all* long A-bar dependencies involve one chain in which there is successive-cyclic movement via SpecCP.[8,9] This goes against a recent proposal by den Dikken (2009a,b), who argues (in part basing himself on earlier work by Rackowski and Richards 2005) that there is no need to assume successive-cyclic movement via SpecCP, and that assuming successive-cyclic movement via SpecCP in fact makes the wrong empirical predictions for languages like Hungarian (see appendix A for discussion of den Dikken's proposal). At this point, it is unclear whether the ban on successive-cyclic movement via SpecCP holds universally.[10]

As for the *trigger* of successive-cyclic movement via SpecCP, I adopt a proposal by Bošković (2007), who argues that the first step of movement in a long-distance A-bar chain (namely movement from the thematic base position to the lower SpecCP position, thereby abstracting away from successive-cyclic movement via the edge of *v*P) is not movement that is triggered by the need to check a (*wh*-)feature. Bošković argues that the driving force behind movement is not an uninterpretable feature (uF) on a probe (as is the 'traditional' minimalist assumption), rather it is an uninterpretable feature (uF) on the goal itself that needs to be checked.[11] Assuming that the uF feature on the goal acts as a probe itself, and assuming that a probe must c-command its goal, the uF feature on the goal triggers movement as it needs to c-command its checker. This

[8] For ease of exposition, I ignore *v*P as a phase (*Phase Theory*, Chomsky 2000 *et passim*), and thus disregard successive-cyclic movement to the edge of *v*P.

[9] Long-distance *wh*-movement is only allowed under *bridge verbs* (cf. Erteschik-Shir 1973), typically verbs of saying (*verba dicendi*), verbs of thinking (*verba putandi*), and verbs of feeling (*verba sentiendi*) – languages may differ from each other with respect to which verbs may and which verbs may not behave as bridge verbs. Complements of non-bridge verbs are islands for extraction, e.g. factive verbs: **de man die ik haat/betreur dat zij uitgenodigd heeft* 'the man who I hate/regret that she has invited'. Most Dutch examples in this chapter feature the bridge verb *denken* 'to think'.

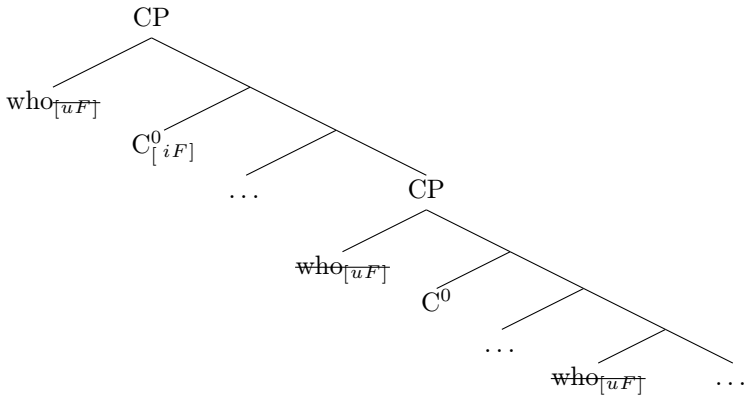
[10] Under any analysis that assumes that movement to SpecCP is terminal (e.g. also Koopman and Sportiche 2008), doubling patterns need to be accounted for by means of multiple chains. The natural question then arises as to what is the relation between these two chains, more specifically, what is the status of the lower clause (that contains the lower chain). It is notoriously difficult to provide an adequate answer to this question; to the best of my knowledge, at this point no satisfactory account exists for Dutch (see appendix A to this chapter for details).

[11] I abstract away from the difference between *uninterpretable* features and *unvalued* features (if any), cf. Pesetsky and Torrego (2007) a.o. for some discussion.

accounts for intermediate movement steps in successive-cyclic *wh*-movement without *look-ahead* in the following way. Suppose we want to build a sentence like (25). Once the first CP is built, uF on *who* indicates that this CP will crash unless *who* is fronted to the phase edge (SpecCP). When the second CP is built, uF on *who* forces it to move to SpecCP of this higher CP in order to c-command its checker: iF on C^0. This is abstractly illustrated in (26).

(25) $[_{CP}$ Who do you think $[_{CP}$ (that) they have called]]?

(26)



It is thus not an EPP feature on the probe of operator movement in the higher CP that forces movement from the base position to the lower SpecCP, but it is the uF on the *wh*-element itself. The first step in long-distance *wh*-movement thus does not take place for (*wh*-)feature checking reasons, as the lower C^0 crucially is not endowed with a *wh*-feature. This is witnessed by the ungrammaticality of (27a), as opposed to (27b). That is to say, the verb *to think*, in contrast to the verb *to wonder*, does not select for a [+wh] complement. It thus cannot be a *wh*-feature on the lower CP in (25) that triggers movement of the *wh*-element to the lower SpecCP.

(27) a. * I think $[_{CP[-wh]}$ who they have called].
b. I wonder $[_{CP[+wh]}$ who they have called].

In short, I take long A-bar dependencies to be derived by successive-cyclic movement via SpecCP, in which movement is considered to consist of the two operations *copy* and *merge* – the Copy Theory of Movement (Chomsky 1993). For convenience I will refer to all chain links – including the highest chain link (commonly referred to as the *head* of the chain) – as *copies*.

2.2.3 Multiple copy spell out

Nunes (2004) assumes that every chain link is computed for linearization in accordance with Kayne's (1994) *Linear Correspondence Axiom* (LCA). Heads as well as lower copies of chains are thus assumed to be subject to the same

principles and mechanisms. According to the LCA a node A precedes a node B if and only if A asymmetrically c-commands B.[12] Under the assumption that two copies of one and the same element count as *identical* for linearization purposes (i.e. they are *non-distinct*), it follows that it is impossible to linearize structures containing identical copies because an element intervening between two copies should simultaneously *precede* and *follow* the same element, which is logically impossible.[13] Deletion of all but one copy is thus a by-product of the requirement to linearize a structure. To account for the observation that in most cases the highest copy of a chain gets pronounced, Nunes (2004) argues that in the standard case, the copy with the most formal features checked gets phonetically realized, i.e. the choice for which copy in a movement chain is pronounced is governed by economy considerations.

Assuming that doubling involves multiple copy spell out, the natural question arises as to how doubling is even possible, i.e. why do doubling structures like (28) and (29) not cause any problems for linearization?

(28) **Wen** glaubt Hans **wen** Jakob gesehen hat?
whom think Hans whom Jakob seen has
'Who does Hans think Jakob saw?' [German, Nunes 2004:39]

(29) **Wêr** tinke jo **wêr**'t Jan wennet?
where think you where-that Jan lives
'Where do you think Jan lives?' [Frisian, Nunes 2004:39]

To account for the doubling data in terms of multiple copy spell out, Nunes assumes that *wh*-pronouns are heads and that heads can undergo *morphological reanalysis* under specific circumstances (cf. the operation *Fusion* in the Distributed Morphology framework). Morphological reanalysis takes two terminal nodes and fuses them together into a single terminal node, as a result of which the number of independent morphemes is reduced. Assuming that the LCA does not apply word internally (Chomsky 1995:337) and assuming that successive-cyclic *wh*-movement (in languages that allow spell out of lower copies) proceeds by adjunction to an intermediate C^0, we can now give an explanation for the well-formedness of doubling structures like (28) and (29). The *wh*-copy in the lower C domain – being head-adjoined to C^0 – undergoes morphological reanalysis with this C^0. The two heads now become a single morpheme, as a result of which the *wh*-copy becomes invisible to the LCA and may thus be spelled out. This is abstractly illustrated in (30).

(30) $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ $[_C$ **pronoun**$_1$ $[_C$ Ø]] ... ~~pronoun$_1$~~ ...]]

[12] In which *asymmetric c-command* is defined as follows: X asymmetrically c-commands Y if and only if X c-commands Y and Y does not c-command X (Kayne 1994:4).

[13] If copies were *distinct* for linearization purposes, spelling out copies within a single movement chain would not be problematic at all. In fact, in such a state of matters, spelling out copies in one and the same movement chain (doubling) would be the *default*. Instead of a *spell out* mechanism, a *deletion* mechanism would then be needed to not overgenerate, and to account for the non-doubling cases (which are most common).

Nunes' proposal regarding the conditions and requirements on the spell out of *wh*-chains is by far the most worked out analysis of multiple copy spell out in the current literature. A great advantage of this theory is that it immediately accounts for the important observation that *wh*-pronouns in doubling constructions are never spelled out in their thematic base position, as illustrated in (31).

(31) * **Wie** denk je <**wie**/ **dat**> ze **wie** geroepen hebben?
who think you who that they who called have
INTENDED: 'Who do you think they have called?'

Recall that in the standard case the copy with the most formal features checked – namely the head of the chain – gets spelled out. In addition to spelling out the head of the chain, a lower copy can only be spelled out when it has undergone morphological reanalysis, thereby becoming invisible to the LCA. Since morphological reanalysis is not possible with the copy of the *wh*-pronoun in base position, but can only target the C head and the copy of the *wh*-pronoun in the CP domain, the fact that copies of *wh*-pronouns can never be spelled out in thematic base position follows from Nunes' analysis.[14]

Moreover, without any additional assumptions, Nunes' analysis accounts for the well known ban on doubling with complex *wh*-phrases, as is illustrated in (32), because only heads are subject to the process of morphological reanalysis.

(32) * **Welke man** denk je **welke man** ik gisteren gezien heb?
which man think you which man I yesterday seen have
INTENDED: 'Which man do you think I have seen yesterday?'

However, Nunes' analysis of multiple copy spell out faces some important problems as well, both theoretically and empirically. First, Nunes' assumption that a pronoun can undergo morphological merger with a C head is hard to reconcile with a phrasal analysis of pronouns – which has been successfully pursued in recent years (e.g. Cardinaletti 1994, Cardinaletti and Starke 1999, Wiltschko 1998, Déchaine and Wiltschko 2002 amongst many others). Second, it should be theoretically impossible for a copy to move any further from a head-adjoined position (cf. Felser 2004:556). Empirically, it is hard to see how Nunes' analysis can account for doubling of prepositional phrases – illustrated in (33) for German and in (34) for Dutch (cf. section 2.7.2) – as PPs are phrases and are thus predicted not to be able to undergo *morphological merger* with the C head.[15]

[14] Alternatively, Thoms (2010) argues that *wh*-copies (of type <et,t>) cannot be interpreted in thematic base positions, because of a semantic type mismatch. If it can be shown that there is a relation between the uninterpretability of a *wh*-copy and the impossibility of spelling out that *wh*-copy, this could lead to an alternative analysis of the impossibility of spelling out *wh*-copies in thematic base positions.

[15] Nunes (2004:169) mentions in a footnote that doubling of prepositional phrases should be accounted for by assuming that the preposition and the *wh*-pronoun first undergo fusion, after which this complex fuses with the C head. This approach runs the risk of overgeneration, predicting all simplex prepositions plus *wh*-pronouns to be able to double. Further research is necessary to check the restrictions (if any) on doubling involving different (simplex) prepositions. See section 2.7.2 for some discussion on doubling involving prepositional phrases.

(33) **Mit wem** glaubst du **mit wem** Hans spricht?
with whom thinks you with whom Hans talks
'With whom do you think Hans is talking?'
[German, Nunes 2004:169]

(34) **Op wie** denk je **op wie** hij verliefd is?
on who think you on who he in love is
'Who do you think he is in love with?'
[colloquial Dutch, MPQ2-B data]

Furthermore, as shown by BKL (2009:23-25), Nunes' analysis makes the wrong empirical predictions as far as *non-identical* doubling is concerned. The remainder of this section summarizes their arguments. Nunes can account for non-identical doubling by assuming a 'big DP analysis' (cf. Uriagereka 1995, Poletto and Pollock 2004, Belletti 2005 and Poletto 2006 a.o.), according to which the two elements in a doubling chain both originate within a single DP and one of them (or both) move(s) out. Under a big DP analysis, the relevant part of a doubling chain involving *wie* and *wat* looks like (35a) or (35b) before PF deletion has taken place.

(35) a. $[_{CP}\ [_{XP}\ \textit{wat wie}] \ldots\ [_{CP}\ [_{XP}\ \textit{wat wie}] \ldots$
b. $[_{CP}\ [_{XP}\ \textit{wie wat}] \ldots\ [_{CP}\ [_{XP}\ \textit{wie wat}] \ldots$

Nunes (2004:26ff.) assumes an operation called *scattered deletion* (cf. Ćavar and Fanselow 1997), which allows deletion to target different constituents within different links in a movement chain. In (35), the higher copy may then for example be spelled out by *wat* whereas the lower copy may be spelled out by *wie*. There are two problems with this analysis of non-identical doubling – and with any analysis that starts out from a big DP approach to doubling more generally. First, the putative big DP never overtly occurs as one constituent (this is certainly unexpected if subextraction from big DP is parallel to *wat voor* split, cf. BKL (2009:20-22) and see section 2.5.5), as illustrated in (36).

(36) a. * **[wat wie]**/ **[wie wat]** denk je dat het gedaan heeft?
what who who what think you that it done has
b. * het meisje **[wat wie]**/ **[wie wat]** ik denk dat het gedaan heeft
the girl what who who what I think that it done has

The observation that the putative big DP never overtly occurs as one constituent is further illustrated by (37), which shows a multiple *wh*-Q. As the pronoun *wie* occupies the sentence initial position, the big DP is expected to be fully spelled out in-situ (in its argument position), *quod non*.

(37) * Wie heeft **[wat wie]**/ **[wie wat]** gezien?
who has what who who what seen
INTENDED: 'Who has seen who?'

Second, the analysis encounters an overgeneration problem. It predicts that the higher copy in (35) may be spelled out as *wie* and the lower copy may be spelled out as *wat*, in contrast to what we find. *Scattered deletion* is thus not restrictive enough. The same holds for a 'big DP plus subextraction' analysis: if subextraction from [$_{XP}$ *wat wie*] or [$_{XP}$ *wie wat*] is possible, a chain in which the higher copy is spelled out as *wie* and the lower copy is spelled out as *wat*, is incorrectly predicted to be grammatical. Additional assumptions are thus required to block subextraction of *wie* in this case.

In the next sections, it will be shown that these problems can be overcome while maintaining the desirable features of Nunes' (2004) analysis.

2.2.4 The analysis of doubling in *wh*-questions: Barbiers, Koeneman, and Lekakou (2009)

BKL start from the assumption that all attested patterns of pronoun doubling in Dutch – repeated here for root *wh*-Qs in (38) – are instances of long-distance movement via SpecCP plus multiple copy spell out à la Nunes (2004) (cf. section 2.2.3).

(38) a. % **Wie** denk je **wie** het gedaan heeft? = (19)
who think you who it done has

b. % **Wie** denk je **die** het gedaan heeft?
who think you RP it done has

c. % **Wat** denk je **wie** het gedaan heeft?
what think you who it done has

d. % **Wat** denk je **die** het gedaan heeft?
what think you RP it done has
'Who do you think has done it?'

(39) a. ?* **Wie** denk je **wat** het gedaan heeft? = (20)
who think you what it done has

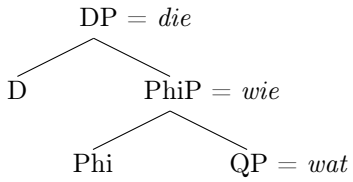
b. * **Die** denk je **wie** het gedaan heeft?
RP think you who it done has

c. * **Die** denk je **wat** het gedaan heeft?
RP think you what it done has

Based on the patterns in (38), and their ungrammatical counterparts in (39), BKL put forward the generalization that in a syntactic movement chain, a higher chain link can never be more specified than a lower chain link (Barbiers 2006). This generalization follows from the following assumptions: (i) pronouns have internal structure and spell out phrases/non-terminals (cf. Weerman and Evers-Vermeul 2002, Neeleman and Szendröi 2007), (ii) syntactic copying can optionally be *partial* (cf. Hiemstra 1986, Cheng 2000), and (iii) PF spell out is all or nothing, i.e. there is no partial spell out at PF (contra the *scattered*

deletion approach, cf. *supra*). The specific structure BKL assume for the Dutch pronouns *die*, *wie* and *wat* is given in (40a). As I will show in detail in section 2.5 that this structure cannot be correct on empirical grounds, I will not go into the argumentation BKL provide to argue for this particular structure.

(40) a.



b. *wie* = *wat* + phi-features

c. *die* = *wie* + definiteness

For identical doubling, BKL follow Nunes (2004) in assuming that doubling is the spell out of multiple copies within one chain (cf. section 2.2.3). This immediately raises the question of how to reconcile Nunes' analysis of identical doubling with the claim that pronouns have internal structure. BKL assume that under certain circumstances non-terminal nodes can undergo morphological merger. Put differently, *morphological reanalysis* may target the C head and the DP/PhiP/QP in SpecCP. Independent evidence for this claim comes from the morphological process of auxiliary contraction as illustrated in (41). The examples in (41a) and (41b) seem to suggest that for auxiliary contraction to be possible, the two terminal nodes need to be in a c-command relation. However, the grammaticality of the sentence in (41c) suggests that the c-command requirement may be satisfied by a *projection* of the terminal node as well.

(41) a. You've got a lot of nerve.

b. * John and you've got a lot in common.

c. Not even you've solved this problem. [BKL 2009:15]

Non-identical doubling, on the other hand, is the result of *partial copying* (what BKL refer to as *partial copying*, I refer to as *subextraction*). Partial copying may target a subpart of the structure in (40), resulting in the spell out of the subextracted element in the higher CP and the spell out of the *full* copy lower down, for reasons of recoverability (notice that *full* spell out of the lower copy is also ensured by the assumption that PF spell out is all or nothing, cf. *supra*).[16] More specifically, starting out with a DP, partial copying may target PhiP, giving rise to the *wie-die* pattern as can be seen in (42a), or it may target QP, giving rise to the *wat-die* pattern as illustrated in (42b). The *wat-wie* pattern is the result of partial copying targeting the QP part of a PhiP, as illustrated in (42c).

[16] It is unclear why subextraction from base position (and thus spell out of the copy in base position) is impossible under this approach. The impossibility of spelling out a copy in base position might be related to the impossibility of interpreting a *wh*-copy in base position (cf. Thoms 2010 and see footnote 14).

(42) a. [$_{CP}$ [PhiP [QP]] … [$_{CP}$ [DP [PhiP [QP]]] … ~~[DP [PhiP [QP]]]~~ …]]
= **wie** = **die**

b. [$_{CP}$ [QP] … [$_{CP}$ [DP [PhiP [QP]]] … ~~[DP [PhiP [QP]]]~~ …]]
= **wat** = **die**

c. [$_{CP}$ [QP] … [$_{CP}$ [PhiP [QP]] … ~~[PhiP [QP]]~~ …]]
= **wat** = **wie**

The ungrammatical doubling patterns in (20)/(39) are cases of full copying and adding structure and features during the course of the derivation, in violation of the *Inclusiveness Condition* (Chomsky 1995:228), which states that outputs cannot contain anything beyond their inputs.[17] This is abstractly illustrated in (43).

(43) a. * [$_{CP}$ [DP [PhiP [QP]]] … [$_{CP}$ [PhiP [QP]] … ~~[PhiP [QP]]~~ …]]
= **die** = **wie**

b. * [$_{CP}$ [DP [PhiP [QP]]] … [$_{CP}$ [QP] … ~~[QP]~~ …]]
= **die** = **wat**

c. * [$_{CP}$ [PhiP [QP]] … [$_{CP}$ [QP] … ~~[QP]~~ …]]
= **wie** = **wat**

It is evident that the analysis of BKL can elegantly account for *all* and *only* the attested patterns of pronominal doubling in long-distance *wh*-Qs. However, it is not directly evident how BKL account for doubling patterns that involve a complex *wh*-phrase, like (44b). Whereas they mention the construction in (44a) and argue that this pattern comes about by subextraction of a subconstituent (QP) from the complex *wh*-phrase *welke man*, the same construction with *wie* in the higher clause is left unmentioned and unexplained. Both constructions in (44) are attested frequently in MPQ1-B (respectively 72% and 63% of the informants accept these sentences), cf. section 2.7.1.

(44) a. % **Wat** denk je **welke man** ik gisteren gezien heb?
what think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'

[17] It is not immediately evident that the ungrammatical doubling patterns are necessarily ruled out by the Inclusiveness Condition. That is to say, several operations could "mask the effect of the Inclusiveness Condition and produce the undesired structures" (BKL 2009:13). For example, the higher copy could enter into an agreement relation with a head in terms of feature copying, as a result of which it becomes more specified than the lower copy. However, besides it being unclear if Spec-Head agreement involves feature *copying*, the matrix SpecCP in root *wh*-Qs is an unlikely location for copying e.g. ϕ-features in case of doubling pattern *wie-wat* (BKL 2009:13–14). Alternatively, as pointed out to me by Marcel den Dikken, it is imaginable that the *lexical array* (or *numeration*) of the higher phase contains a D^0 with which the highest copy of a PhiP merges, thereby deriving the ungrammatical doubling pattern *die-wie*. Such an operation can be ruled out by appealing to the *extension condition* (Chomsky 1995:190–191,327–329), which requires syntactic operations to extend the syntactic tree at the root. In addition, the rationale behind the presence of a D^0 in the higher phase of a long-distance root *wh*-Q is unclear.

b. % **Wie** denk je **welke man** ik gisteren gezien heb?
who think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'

More importantly, BKL's analysis of doubling in long-distance *wh*-Qs does not carry over to RCs. On the basis of the then available data, BKL (2009:3) assumed that the doubling pattern *die-wie* was not attested in RCs, as illustrated in (45b). The absence of this doubling pattern is predicted by their analysis of doubling in long-distance *wh*-Qs – according to which *die-wie* is analyzed as a violation of *Inclusiveness* (cf. (43a)) – and thus seems to suggest that their analysis of doubling is not restricted to *wh*-Qs.

(45) a. Dit is de man **wie** ik denk **die** Jan gezien heeft.
this is the man who I think RP Jan seen has
'This is the man I think Jan has seen.' [Drenthe Dutch]

b. * Dit is de man **die** ik denk **wie** Jan gezien heeft.
this is the man RP I think who Jan seen has
[BKL 2009:3]

However, as I will show in the next section, closer empirical investigation reveals that sentences like (45b) are in fact attested in the Dutch speaking language area, i.e. BKL incorrectly assigned a * to (45b) – an empirical observation that cannot be accounted for by the analysis of BKL in its present form.

2.3 Doubling in long-distance relative clauses

Relative clauses in Standard Dutch are head-initial, they always need to be introduced by a relative pronoun, and the relativized item leaves a *gap* at the extraction site (not a resumptive pronoun), cf. chapter 1. Standard Dutch restrictive RCs thus obey the format in (46).

(46) $[_{matrix\ clause}$ ... RC head $[_{RC}$ *(relative pronoun) ... *gap* ...]]
Standard Dutch restrictive relative clause

The most prominent analyses of RCs in current literature – the *head raising analysis* (cf. Vergnaud 1974, Kayne 1994, Bianchi 1999, 2000, and de Vries 2002 a.o.) and the *matching analysis* (cf. Chomsky 1965, Munn 1994, Citko 2001, and Salzmann 2006 a.o.) – assume that RCs are derived by A-bar movement of the relative pronoun together with the RC head, i.e. there is an (additional) RC internal representation of the RC head. It is immaterial at this point whether or not the syntax of RCs involves *raising* of the RC head. For ease of exposition, in this chapter, I will therefore simply assume that only the relative pronoun undergoes A-bar movement in RCs, just like the *wh*-pronoun does in *wh*-Qs. More specifically, in this chapter, I assume a traditional *Head External Analysis* of RCs, according to which the RC head is base-generated outside of the relative

clause CP, and inside the RC a relative pronoun moves to the left periphery (where it is linked to the RC head by means of predication; cf. Quine 1960, Chomsky 1977, Smits 1988, Borsley 1997 a.o.). In chapter 3, I make a case for (a particular implementation of) the HEA of RCs.

In Standard Dutch long-distance restrictive RCs, the RC itself is introduced by a relative pronoun and the embedded finite declarative clause is introduced by the finite declarative complementizer *dat* 'that'. Relative pronouns in restrictive RCs in Standard Dutch are (partly) taken from the *d*-pronoun paradigm and agree in gender and number with their antecedent (i.e. the RC head).[18,19] Modern Standard Dutch has a two-way gender contrast within the relative pronoun paradigm, namely *neuter* and *common* (=*non-neuter*) gender (cf. Audring 2009):[20] *die* is the common gender relative pronoun, and *dat* is the neuter gender relative pronoun. The latter is identical in form to the Dutch finite declarative complementizer (see chapter 4 for discussion on the (relative) pronoun/complementizer alternation).

The examples in (47a) and (47b) show subject and object relativization with the common gender antecedent *man* 'man' that requires the relative pronoun *die* in Standard Dutch. All examples in this section feature the common gender antecedent *man*.

(47) a. Dat is de man **die** ik denk **dat** het verhaal verteld heeft.
that is the man RP I think that the story told has
'That is the man who I think told the story.'

b. Dat is de man **die** ik denk **dat** ze geroepen hebben.
that is the man RP I think that they called have
'That is the man who I think they have called.'

As can be seen in (47) – and as could already be seen from the obligatory presence of *dat* in the lower clause of long-distance subject *wh*-Qs in Standard Dutch (cf. section 2.2.1) – in contrast to languages like English, Dutch does not show a *that*-t effect (e.g. Perlmutter 1971, Chomsky and Lasnik 1977). The complementizer is always obligatorily present, independently of subject or object extraction.[21] More generally put, the declarative complementizer *dat* in Dutch cannot be left out in embedded declarative clauses, and no variety of

[18] In this chapter, I abstract away as much as possible from number, i.e. the singular/plural distinction, as taking this distinction into account would unneccesarily complicate matters.

[19] See van Kampen (2007, 2010) amongst others for a detailed overview of the Dutch relative pronoun system.

[20] Only the personal pronoun system in modern Standard Dutch has retained three genders (*masculine*, *feminine* and *neuter*).

[21] Although Dutch does not have a *that*-t effect that manifests itself as the absence of complementizer *dat* 'that' with subject extraction from a finite embedded clause, the extractability of the subject across the finite complementizer seems to be influenced by other factors (see den Dikken 2007 for an overview). For example, for some Dutch speakers and/or in some varieties of Dutch, insertion of the expletive *er* 'there' (*er*-insertion) improves or enables subject extraction (cf. Haegeman 2004a, den Dikken 2007 a.o.).

Dutch allows zero-relativization (Dekkers 1999, Boef 2008b).[22,23] The format that Standard Dutch long-distance restrictive RCs obey is given in (48).

(48) ... RC head [$_{RC}$ *(RP) ... [*(complementizer) ... *gap* ...]]
Standard Dutch long-distance restrictive relative clause

It is worth mentioning here that in addition to the long-distance RC constructions in (47), Dutch has an alternative way of forming long-distance RCs, which is illustrated in (49). In (49), the relative pronoun in the higher clause is governed by a preposition – the relative pronoun is taken from the *w*-pronoun paradigm because *d*-pronouns generally do not occur as the object of a preposition (unless with strong emphasis), cf. section 4.2.2 – and a resumptive pronoun appears at the extraction site (either a personal or a demonstrative pronoun). This construction is semantically equivalent to the construction in (47) and is referred to as *resumptive prolepsis* (Salzmann 2006).

[22]In infinitival RCs, complementizer *om* is obligatorily present as well (ia), whereas it may be optional when introducing an infinitival sentential complement (ib).

(i) a. een leuk meisje *(om) mee uit te gaan
a nice girl to with out to go
'a nice girl to go out with'

b. ik beloof (om) mee uit te gaan
I promise to with out to go
'I promise to go out'

[23]An (apparent) counterexample to the claim that embedded declarative clauses are always introduced by complementizer *dat* is provided by the data in (i) from the SAND corpus: some dialects allow embedded V2 (with certain verbs) without the complementizer being present (22 out of 75 locations, cf. DynaSAND) – this phenomenon is quite common in German. Notice that we can maintain the generalization that in sentences with subordinate clause word order (V-final), complementizer *dat* always needs to be overtly present, as in (ii) (there are only two exceptions in the SAND corpus).

(i) Ik zei nog tegen haar: ik denk hij is weg.
I said just to her I think he is gone
(ii) Ik zei nog tegen haar: ik denk *(dat) hij weg is.

Even though the SAND interviewers were instructed to pronounce the part of the sentence after the colon in (i) as one intonation phrase, it is not completely clear that the sentence in (i) does not involve *recursive* or *double* direct speech (Hans Bennis p.c.). If it would, (i) would not constitute a counterexample to the generalization after all. As suggested to me by Sjef Barbiers, whether or not we are dealing with double direct speech could be tested by the sentence in (iii). Since variable binding is usually impossible across a sentence boundary, if informants accept (iii) with a bound reading (i.e. *hij* is bound by *iedereen*), this would indicate that there is no sentence boundary (and thus no double direct speech). In that case, it would really constitute a counterexample to the claim that embedded declarative clauses are always introduced by complementizer *dat*.

(iii) Ik zei nog tegen haar: iedereen denkt hij is weg.
I said just to her everybody thinks he is gone

(49) a. de man **van wie** ik denk dat **hij/die** het verhaal verteld heeft
the man of who I think that he/DEM the story told has
'the man who I think told the story'

b. de man **van wie** ik denk dat ze **hem/die** geroepen hebben
the man of who I think that they him/DEM called have
'the man who I think they have called' [Standard Dutch]

In Dutch it is independently possible to invert the *w*-pronoun and the preposition, in which case the pronoun turns into the R-pronoun *waar* (cf. van Riemsdijk 1978). The sentences in (50) are thus the equivalent of the sentences in (49).[24] In fact, in the SAND corpus, long-distance RCs with *waarvan* 'whereof' were attested more frequently than their counterpart with *van wie* 'of who'.

(50) a. de man **waarvan** ik denk dat **hij/die** het verhaal verteld heeft
the man whereof I think that he/DEM the story told has
'the man who I think told the story'

b. de man **waarvan** ik denk dat ze **hem/die** geroepen hebben
the man whereof I think that they him/DEM called have
'the man who I think they have called' [Standard Dutch]

Salzmann (2006) argues that the constructions in (49)-(50) involve *indirect* dependencies, i.e. the different syntactic objects (the boldfaced elements in (49)-(50)) are not part of the same movement chain. In contrast, I take the constructions in (47) to involve *direct* A-bar dependencies, and thus to differ considerably from the resumptive prolepsis construction. Therefore, in the remainder of this thesis I will abstract away from the latter construction, and from *resumption* in long-distance A-bar dependencies more generally.[25,26]

[24] According to the ANS (Haeseryn et al. 1997:344), *van wie* 'of who' is used when the antecedent refers to a person, whereas *waarvan* 'whereof' is used when the antecedent refers to a thing/matter. The use of the construction in (50) as an alternative for the construction in (49) – i.e. *waarvan* 'whereof' used with a human antecedent – is mostly said to occur in *informal language*.

[25] See Salzmann (2006) for a detailed overview of the properties and an analysis of the resumptive prolepsis construction, and see McCloskey (2006) and references cited therein for an overview of resumption in (long-distance) A-bar dependencies.

[26] Although the long-distance relativization constructions in (47) are 'officially' Standard Dutch (cf. ANS, Haeseryn et al. 1997:1304), not all Dutch speakers accept these sentences (and some speakers prefer the *resumptive prolepsis* construction in (49)-(50)). According to the ANS, the construction in (47) is not very common, and it occurs mainly in written language. A similar observation was already made by Brachin (1973, 1974) for long-distance *subject* relativization, as in (47a). The MPQ1-B data furthermore show that only a bit more than 60% of the informants accept the sentences in (47). Even so, I will continue to refer to the construction in (47) as the Standard Dutch variant.

In colloquial Dutch long-distance RCs – just as in colloquial Dutch long-distance (embedded) *wh*-Qs – both the higher clause and the lower clause can be introduced by a pronoun, the form of which is dependent on the antecedent (cf. section 2.5.2).[27] This gives rise to *identical* doubling, as illustrated in (51) for RCs with the common gender human antecedent *man* 'man'.

(51) a. % Dat is de man [$_{RC}$ **die** ik denk [**die** het gedaan heeft]].
that is the man RP I think RP it done has
'That is the man who I think has done it.'

b. % Dat is de man [$_{RC}$ **wie** ik denk [**wie** het gedaan heeft]].
that is the man who I think who it done has
'That is the man who I think has done it.'

In addition to *identical* doubling with pronouns *die* and *wie* in long-distance RCs – and long A-bar dependencies more generally – in colloquial Dutch, the RC itself, as well as the lower clause, can be introduced by *non-identical* pronouns, as illustrated in (52). In (52a) the higher pronoun is *wie* and the lower pronoun is *die* (*wie-die*), whereas (52b) shows exactly the opposite pattern (*die-wie*).

(52) a. % Dat is de man [$_{RC}$ **wie** ik denk [**die** het gedaan heeft]].
that is the man who I think RP it done has
'That is the man who I think has done it.'

b. % Dat is de man [$_{RC}$ **die** ik denk [**wie** het gedaan heeft]].
that is the man RP I think who it done has
'That is the man who I think has done it.'

Identical doubling of pronoun *wat* in RCs with a common gender human antecedent like *man* 'man' in (51c) was not explicitly tested in the MPQs, but it occurs very marginally in the SAND corpus (as a spontaneous translation of (47)), namely in Rijckholt and Vaals, in the southern part of Dutch Limburg. Similarly, the MPQ1-B data show that non-identical doubling involving pronoun *wat* in RCs with a common gender antecedent is not or very marginally attested, as illustrated in (54).

(53) ?* Dat is de man [$_{RC}$ **wat** ik denk [**wat** het gedaan heeft]].
that is the man what I think what it done has
INTENDED: 'That is the man who I think has done it.'
[SAND data]

[27] More than half of the MPQ1-A informants make use of *wie* 'who' as a relative pronoun in restrictive RCs that are headed by the common gender human noun *man* 'man' (as opposed to or in addition to *die*). This is illustrated in (i).

(i) % Dat is de man [$_{RC}$ **wie** het gedaan heeft].
that is the man who it done has
'That is the man who has done it.' [279/452=62%, MPQ1-A data]

(54) a. ?* Dat is de man [$_{RC}$ **wat** ik denk [**wie** het gedaan heeft]].
that is the man what I think who it done has

b. ?* Dat is de man [$_{RC}$ **wie** ik denk [**wat** het gedaan heeft]].
that is the man what I think what it done has

c. ?* Dat is de man [$_{RC}$ **wat** ik denk [**die** het gedaan heeft]].
that is the man what I think RP it done has

d. ?* Dat is de man [$_{RC}$ **die** ik denk [**wat** het gedaan heeft]].
that is the man RP I think what it done has
INTENDED: 'That is the man who I think has done it.'

[MPQ1-B data]

The reason for the non-occurrence (or the very restricted distribution) of (53) and (54) is most likely related to the fact that *wat* as a relative pronoun hardly ever occurs with common gender antecedents like *man* 'man' in the first place. It occurs frequently with neuter antecedents like *boek* 'book' or *meisje* 'girl', cf. van Kampen (2007, 2010) and see section 2.5.2. For this reason, in the following sections, I will abstract away as much as possible from all (non-) identical doubling involving *wat* in RCs, but see section 2.6. Rather, I will focus on doubling in RCs involving *die* and *wie*.

The geographic distribution of the doubling patterns in long-distance RCs with the antecedent *man* 'man' as attested in MPQ1-B is illustrated on map 2.2. Just as on map 2.1, the numbers in between the first parentheses after the given doubling pattern give the number of *attestations* of that doubling pattern out of a total of 333 attestations, and the number in between the second parentheses gives the number of *locations* in which the given doubling pattern is attested. The map shows that in most locations more than one doubling pattern is attested. Recall from section 1.3.2 that there can be more than one informant within a single location, as a result of which the attested combinations of doubling patterns represent the attestation of these combinations per *location* and not necessarily per *informant*.

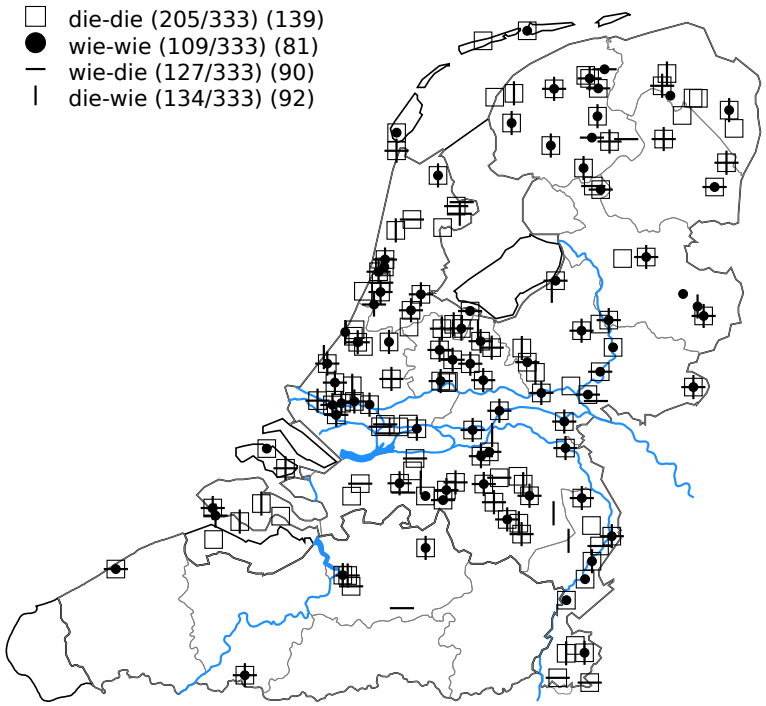


Figure 2.2: Doubling in subject RCs with antecedent *man* (MPQ1-B data)

The observation that doubling pattern *die-wie* in (52b) is attested in long-distance RCs is highly unexpected under BKL's analysis of the doubling patterns in *wh*-Qs, as it violates the *Inclusiveness Condition* (cf. section 2.2.4). Put differently, following the null hypothesis that pronouns in RCs and pronouns in *wh*-Qs are the same elements with the same properties, we are left with two options. Either we take the existence of (52b) to indicate that doubling in *wh*-Qs is different from doubling in RCs, or we take doubling in *wh*-Qs to be the same phenomenon as doubling in RCs, the differences between them being accounted for in another fashion. I will argue for the second option and propose some modifications to the theory of BKL – most prominently, a different structure and a different feature specification of the relevant pronouns – that makes it compatible with doubling in both *wh*-Qs and RCs.

Having established the empirical basis for this chapter, a short digression is in order to point out a complicating factor for determining the nature of the doubled elements in RCs. More specifically, before I give a summary of the doubling patterns in long-distance A-bar dependencies and present my analysis of these facts, the following section focuses on the status of *die* in varieties of Dutch that show subject/object asymmetries in long-distance RCs.

2.3.1 A note on special *die* in relative clauses

Until now, no or very little attention has been paid to the difference between subject and object extraction. It is, however, well known that there are dialects of Dutch that show a subject/object asymmetry in (long-distance) relativization, i.e. the pattern that is used for subject relativization is different from the pattern that is used for object relativization (cf. Haegeman 1983, Bennis and Haegeman 1984, Boef 2008a,b, 2012b). The s/o asymmetry in long-distance RCs that I want to focus on here is illustrated in (55) and (56); these data are taken from the DynaSAND (Barbiers et al. 2006).

(55) a. Da s de vent **da** k peizen **die** da graptje verteld eet.
that is the man that I think *die* that joke told has
'That is the man who I think told that joke.'

b. Da s de vent **da** k peizen **da-n** ze geroepen en.
that is the man that I think that-3PL they called have
'That is the man who I think they have called'.

[Brugge Dutch, West-Flemish]

(56) a. Da s de man **die** k peize **die** t verhaal verteld ee.
that is the man RP I think *die* the story told has
'That is the man who I think told the story.'

b. Da s de man **die** k peize **da-n** ze geropen en.
that is the man RP I think that-3PL they called have
'That is the man who I think they have called.'

[Gent Dutch, East-Flemish]

Although the element that introduces the higher clause differs per location – *dat* in West-Flemish (55) and *die* in East-Flemish (56) – the pattern that is found in the lower clause of long-distance relativization is the same across the southern Dutch varieties: *die* with subject extraction and *dat* with object extraction.[28] I will refer to the element *die* that is found in the most deeply embedded clause of (long) subject RCs in dialects that show a s/o asymmetry with 'special *die*'. I borrow this term 'special *die*' from Koopman and Sportiche (2008), but whereas they use it to refer to *die* in the lower clause of both subject and object extraction, I use 'special *die*' in a more restricted way, namely only in the context of *subject* extraction, and only in the context of *relative clauses*. For Dutch it can be shown that the distribution of special *die* across construction types is restricted: it occurs almost certainly only in RCs, but not in other long A-bar dependencies, cf. section 4.5. The definition of special *die* is given in (57).

[28] In varieties that show a s/o asymmetry, the complementizer can have the form *dat* or *da* (and it may have an additional agreement suffix), cf. (55)-(56). In the main text, I will simply use *dat* for all form variants.

(57) **special *die***: *die* in the left periphery of the most deeply embedded clause in (long) subject RCs in dialects that show a s/o asymmetry

In section 4.5, I will argue in detail that special *die* is not a relative pronoun but an (agreeing variant of the) complementizer (cf. Rizzi 1990 for French *qui*). Put differently, the element *die* in the lower clause of the a-sentences in (55) and (56), I argue is not a pronoun. Consequently, the *die-die* pattern in (56a) is not an instance of doubling of the pronoun *die*, rather we are dealing with an instance of relative pronoun *die* in the higher clause, and an instance of complementizer *die* in the lower clause. How do we know then that the *subject* RCs in which the lower clause is introduced by *die* as introduced in the previous section, do not involve special *die*? In other words, how can we be sure that we have been dealing with real instances of doubling of a pronoun, instead of a combination of a pronoun in the higher clause and an (agreeing variant of the) complementizer *die* in the lower clause?

The SAND data show that the geographic distribution of special *die* is restricted (cf. also the maps in Boef 2008b, 2012b). It almost exclusively occurs in West- and East-Flanders; only the pattern in (56) is attested sporadically in the east of the Netherlands. Since the MPQ data are mainly from the Netherlands (cf. section 1.3.2), and special *die* is not or very marginally attested within this corpus, we can safely conclude that the doubling patterns involving *die* as presented in the previous section do not involve special *die*. Therefore, I assume that all cases of doubling in RCs as presented in this chapter involve true doubling of a pronoun, and I postpone the discussion about special *die* to chapter 4.

Having determined that the patterns of doubling as laid out in the previous section do not involve special *die*, but are real instances of doubling of a relative pronoun, the next part of this chapter is devoted to the analysis of these doubling configurations, and doubling of pronouns in long-distance A-bar dependencies more generally. But first, the next section summarizes the attested patterns of doubling in *wh*-Qs and RCs.

2.4 Data summary

The table below summarizes the attested patterns of doubling – semantically equivalent to their non-doubling counterparts – in long-distance restrictive RCs with a common gender human antecedent (*man* 'man'), and the patterns of doubling in long-distance (embedded) *wh*-Qs that question a person (*wie* 'who').

Table 2.1: Doubling patterns in RCs and *wh*-Qs in colloquial Dutch

	restrictive relative clauses with antecedent *man* 'man' (common gender)	embedded *wh*-questions that question a person (*wie* 'who')
die-die	+	*
wie-wie	+	+
wat-wat	?*	*
wie-die	+	+
die-wie	+	*
wat-wie	?*	+
wie-wat	?*	?*
wat-die	?*	+
die-wat	?*	*
+ = attested in MPQ1-B and/or SAND corpus * = not attested in MPQ1-B and/or SAND corpus ?* = very marginally attested in MPQ1-B and/or SAND corpus[29]		

2.5 The proposal

Assuming that pronouns have internal structure, in this section, I will argue that doubling in long-distance A-bar dependencies is the result of multiple copy spell out (doubling involving only *die* and/or *wie*) or of subextraction of part of the pronoun, namely the operator in its specifier position (non-identical doubling involving *wat* in the higher clause). This operator is subextracted from the lower SpecCP position and spelled out in its final landing site, the higher SpecCP. The copy of the pronoun that is left behind by subextraction is spelled out as well, for reasons of recoverability and in order to overcome an otherwise illicit step in the derivation (*rescue by PF spell out*). The subextracted operator will be spelled out as *wat* by default – *wat* being the most underspecified A-bar pronoun in Dutch.

I first introduce the assumptions I adopt about the internal structure of A-bar pronouns in section 2.5.1, after which I discuss in detail the feature specifications of the relevant A-bar pronouns in Dutch in section 2.5.2. I then proceed with the analysis by making explicit my assumptions about lexicalization and subextraction in sections 2.5.3 and 2.5.4 respectively. Section 2.5.5 discusses the nature of the mechanism *rescue by PF spell out*, and finally, section 2.5.6 gives a summary of all the attested pronominal doubling patterns and their analysis.

[29]Recall from section 1.3.2 that doubling patterns with a ?* grammaticality judgment are considered ungrammatical.

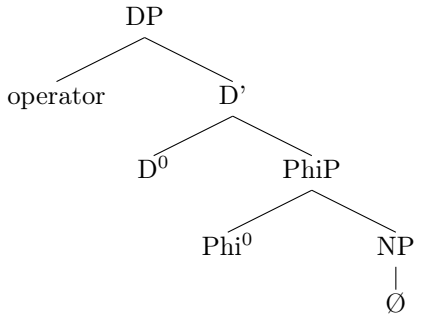
2.5.1 The internal structure of A-bar pronouns

In this section, I make explicit my assumptions about the internal structure of interrogative and relative pronouns. I follow a large body of literature and assume that pronouns have internal structure (e.g. Cardinaletti 1994, Ritter 1995, Noguchi 1997, Wiltschko 1998, Cardinaletti and Starke 1999, Koopman 1999, Harley and Ritter 2002, Déchaine and Wiltschko 2002, Wiltschko 2002, Rooryck 2003, van Koppen 2005), and that pronouns spell out phrases/non-terminals (cf. Weerman and Evers-Vermeul 2002, Neeleman and Szendröi 2007, and BKL a.o.).

On both empirical and theoretical grounds it seems reasonable to assume that (A-bar) pronouns are morphologically and syntactically complex. It is evident that there are different categories of pronouns, and that differences between such categories are reflected both syntactically and morphologically. For example, *strong* personal pronouns behave like full DPs, whereas *weak* or *clitic* pronouns do not behave like full DPs and are often morphologically reduced when compared to their strong counterparts.[30] Such differences can be captured by taking the external syntax of pronouns to be a reflection of their categorial status and internal syntax. A different categorial status – e.g. DP or NP – can account for differences in e.g. syntactic distribution and binding-theoretic status. Furthermore, the fact that clitic or weak pronouns are often *morphologically reduced* can then be taken to be the immediate result of the different syntactic structure they have (i.e. a *reduced* or '*deficient*' syntax, cf. Cardinaletti and Starke 1999).

The structure of A-bar pronouns that I assume is given in (58). More specifically, (58) represents the abstract syntactic structure of A-bar pronouns that feeds into morphology (cf. section 1.2.1).

(58) The structure of A-bar pronouns



Taking A-bar pronouns to be DPs is compatible with the fact that in RCs, the gap inside the RC – where the relative pronoun arguably originates (under a

[30]Most of the literature that argues in favor of a phrasal analysis of pronouns has been directed towards the internal structure of *personal pronouns*, and less towards A-bar pronouns like relative and interrogative pronouns.

Head External Analysis of RCs) – acts as a DP (see e.g. Borsley 1997 and chapter 3).[31] The fact that *die*, *wie* and *wat* all may function as relative pronouns thus indicates that they are indeed DPs. Further evidence in favor of the claim that A-bar pronouns are DPs, comes from *binding*. That is, if A-bar pronouns are pro-DPs in the sense of Déchaine and Wiltschko (2002), they should function as R-expressions with respect to binding. This prediction is borne out, as illustrated in (59) for the A-bar pronoun *die*: *die* is subject to Condition C (59a), and it does not allow for a bound variable interpretation (59b).[32,33]

(59) a. Jan$_i$ denkt dat waarschijnlijk die$_{*i/j}$ de wedstrijd zal winnen.
Jan thinks that probably DEM the game will win
'Jan thinks that probably he will win the game.'

b. Iedere jongen$_i$ denkt dat de vrouw die$_{*i/j}$ aantrekkelijk vindt.
every boy thinks that that woman DEM attractive finds
'Every boy thinks that that woman finds him attractive.'
[Standard Dutch, Corver and van Koppen 2008:10]

Suggestive evidence in favor of the DP status of *w*-pronouns in Dutch comes from the *categorial matching effect* in free relative clauses (FRCs, cf. amongst others Groos and van Riemsdijk 1981, van Riemsdijk 2006 and references cited therein). This effect requires the categorial status of the *wh*-phrase in the left periphery of the FRC to be identical to the categorial status of the whole FRC as required by the matrix clause. This is illustrated by the paradigm in (60). The adjective *verliefd* 'in love' selects for a PP and only in case the *wh*-phrase introducing the FRC is a PP is the structure grammatical: (60a) vs. (60c). Similarly, the verb *kussen* 'to kiss' selects for a DP and only when the *wh*-phrase in the left periphery of the FRC is a DP is the structure grammatical:

[31] In section 2.5.2.4 (footnote 57) I show – following BKL – that the grammaticality of the equivalent of sentences like *The men that there were in the garden were all diplomats* (cf. also section 3.2.1) is not a solid argument against the definiteness and DP status of relative pronouns (in particular pronoun *die*).

[32] As replacing the *d*-pronoun *die* with the *w*-pronoun *wie* in (59) would lead to ungrammaticality, the binding theoretic status of *w*-pronouns cannot be tested.

[33] Jacqueline van Kampen reminds me that *d*-pronouns differ from R-expressions in that the latter need to be free both inside and outside the CP they occur in, whereas *d*-pronouns can – and in fact *need* to – have an antecedent (in the linguistic or extra-linguistic context). The important point is that *d*-pronouns cannot have an antecedent in the CP that they have scope over, cf. Hoekstra (1999): *d*-pronouns must be free in the domain of the matrix sentence. However, the requirement that a *d*-pronoun cannot be bound by a c-commanding antecedent in the matrix sentence appears to be able to be overridden in case of an ambiguous sentence: a *d*-pronoun signals topic-shift and disambiguates the sentence (Marcel den Dikken, p.c.). This is illustrated in (i).

(i) a. Jan$_i$ vertelt Piet$_k$ dat hij$_{i/k}$ het boek moet lezen.
Jan tells Piet that he the book must read
'Jan tells Piet that he (Jan/Piet) must read the book.'

b. Jan$_i$ vertelt Piet$_k$ dat die/deze$_{*i/k}$ het boek moet lezen.
Jan tells Piet that DEM the book must read
'Jan tells Piet that he (Piet) must read the book.'

(60b) vs. (60d). In the sentences in (61), the FRC as a whole acts as a DP, i.e. it occurs in a position otherwise restricted to a DP argument. Following the logic above, this means that the *wh*-phrase introducing the FRC must be a DP as well.

(60) a. Jan is verliefd [$_{PP}$ [$_{PP}$ op wie] Kees verliefd is t_{PP}].
Jan is in love on who Kees in love is
'Jan is in love with who(ever) Kees is in love.'

b. Jan wil kussen [$_{DP}$ [$_{DP}$ wie] Kees kust t_{DP}].
Jan wants kiss who Kees kisses
'Jan wants to kiss who(ever) Kees kisses.'

c. * Jan is verliefd [$_{PP}$ [$_{DP}$ wie] Kees kust t_{DP}].
Jan is in love who Kees kisses

d. ?* Jan wil kussen [$_{DP}$ [$_{PP}$ op wie] Kees verliefd is t_{PP}].
Jan wants kiss on who Kees in love is

(61) a. Ik eet [$_{DP}$ wat jij eet].
I eat what you eat
'I eat what(ever) you eat.'

b. Jan interviewt [$_{DP}$ wie Kees interviewt].
Jan interviews who Kees interviews
'Jan interviews who(ever) Kees interviews.'

It is the DP layer of A-bar pronouns in which (in)definiteness is expressed. A-bar pronouns furthermore have a PhiP layer in which phi-features are expressed, and they contain an empty NP that provides the range for the operator-variable chain, cf. Wiltschko (1998).[34,35,36] Moreover, A-bar pronouns contain an operator that is the driving force behind movement to the left periphery; in terms of the analysis of Bošković 2007 (cf. section 2.2.2), this means that the uF feature that triggers successive-cyclic movement is located on the operator. This operator is located in the SpecDP position, cf. Szabolcsi (1994) a.o., who argues that the specifier of DP is an operator position. Notice that this fits in perfectly with the often noted parallelism between DP and CP (cf. Szabolcsi 1987, 1994, Cardinaletti and Starke 1999, Haegeman and Ürögdi 2010 amongst many others), as SpecCP is the designated position for operator movement in the clausal domain.

[34] In RCs, the value of the range is provided by the RC head (cf. chapter 4).

[35] The presence of an empty NP in A-bar pronouns is required to exclude the use of personal pronouns (that do not contain an NP/range) in RCs (and *wh*-Qs), cf. Wiltschko (1998) and see section 4.2.

[36] The structure proposed here differs from a proposal made for German by Rett (2006), according to which *wh*-pronouns are non-quantificational *wh*-phrases *without* an NP complement. An argument in favor of that claim comes from the fact that in German *wh*-phrases without an NP complement are in complementary distribution with *d*-pronouns that have an NP complement. However, in Dutch this argument does not stand as *d*-pronouns are not in complementary distribution with *wh*-pronouns – as can for example be seen in the doubling cases, where *die* and *wie* are often interchangeable.

In the following sections, I will argue that pronouns *die*, *wie* and *wat* (and *dat*) may lexicalize the full DP, but that in addition, *wat* – being the most underspecified A-bar pronoun in Dutch – may lexicalize the operator when subextracted.

2.5.2 The feature specification of Dutch A-bar pronouns

This section provides the feature specifications of the relevant Dutch A-bar pronouns: *die*, *wie*, and *wat*. For the sake of completeness, and in light of the doubling patterns in RCs with a neuter gender antecedent (cf. section 2.6.1), I will also take into account (the feature specification of) the Dutch A-bar pronoun *dat*.[37] As became clear in the previous sections, these pronouns may appear in different syntactic configurations (the higher and lower left periphery in long-distance *wh*-Qs and RCs). Whereas there exists a reasonable amount of literature exploring the nature of pronouns (e.g. their internal structure and the interaction with their external syntax, cf. Cardinaletti 1994, Postma 1994, Cardinaletti and Starke 1999, Koopman 1999 a.o.), little attention has been paid to what I will refer to as *multipurpose pronouns*: pronouns that can have more than one function and may appear in more than one syntactic configuration. I start from the hypothesis that multipurpose pronouns are morphosyntactically *underspecified* (cf. Rooryck 2003, Postma 1994), as a result of which they are able to appear in different configurations. See chapter 4 for discussion.

In this section, I give a detailed overview of the different uses, functions and properties of the pronouns under discussion, and I provide a feature specification for them that is compatible with all their uses. Variation between speakers in the use of A-bar pronouns, I argue, is not the result of different feature specifications of the relevant A-bar pronouns, i.e. the feature specifications of the elements involved are similar in Standard Dutch and colloquial Dutch. Rather, I will argue that variation between speakers is the result of different lexicalization preferences: spelling out syntactic gender and/or spelling out semantic animacy (cf. section 2.5.3). The next table shows the different functions that the relevant Dutch A-bar pronouns can have (*d*-pronouns: *die* and *dat* and *w*-pronouns: *wie* and *wat*).

[37] The tables in this section – most prominently tables 2.4, 2.5, 2.6, and 2.7 – are compiled from several data sources: the SAND corpus (Barbiers et al. 2005, 2008a), the MPQ data, and several other studies, like Haeseryn et al. (1997), Audring (2009) and van Kampen (2007, 2010). These data mostly concern Standard Dutch and colloquial Dutch; the text contains only scattered, non-systematic remarks about dialectal variation.

Table 2.2: Functions of Dutch A-bar pronouns *die*, *dat*, *wie* and *wat*

	d-pronouns		*w*-pronouns	
	die 'that.C'	*dat* 'that.N'	*wie* 'who'	*wat* 'what'
determiner/nominal modifier	+	+	−	+
demonstrative pronoun	+	+	−	−
relative pronoun	+	+	+	+
interrogative pronoun	−	−	+	+
exclamative marker	−	−	−	+
indefinite pronoun	−	−	−	+
resumptive pronoun	+	+	−	−
complementizer	–[38]	+	−	−

Before I turn to the analysis and feature specification of these elements, in the next section I want to briefly lay out my assumptions regarding the syntactic representation of features in the grammar.

2.5.2.1 On the representation of morphosyntactic features

I assume that morphosyntactic features are represented in syntax by means of an *attribute-value* structure [x:y], where x is the *attribute* and represents the type of feature involved (for example *person*), and y represents the *value* of the feature (for example *first*).[39] Instead of assuming that each attribute needs a value – i.e. each interpretation of a feature is mapped onto an attribute-value pair: *obligatory specification* – I assume that values may be absent, indicating *underspecification* of a given feature (cf. Schoorlemmer 2009 a.o.). When a particular lexical item has an underspecified feature representation, i.e. it lacks a value for a particular attribute, it acts as a variable in the sense that it is flexible in its combinatorial possibilities with other lexical items. However, when the underspecified lexical element occurs on its own, it will get the *default* interpretation. For example, when a lexical element is underspecified for number, it can combine with singular as well as plural elements, but when it occurs on its own it will be interpreted as singular by default.

The representation of the (for present purposes) relevant morphosyntactic features – most prominently **gender** – is given in (62). I also include the semantic feature **animacy** here (62d), as animacy influences pronoun choice in

[38] As mentioned in section 2.3.1 (and see section 4.5 for more details), in some southern Dutch dialects, *die* may function as a complementizer as well. Notice that this is not unexpected when looking at table 2.2: the counterpart of *die*, namely *dat*, functions as a complementizer and they both have exactly the same range of other functions.

[39] Cf. Chomsky (2001), Cheng and Rooryck (2000), Rooryck (2000, 2003), Pesetsky and Torrego (2007), Schoorlemmer (2009) amongst others. In the HPSG framework (*Head-Driven Phrase Structure Grammar*, cf. Pollard and Sag 1994), attribute-value matrices have been widely used for a long time.

RCs and *wh*-Qs (cf. *infra*).[40,41,42] The default values of these features are given in (62e).

(62) a. [referentiality: definite] = definite
[referentiality:] = underspecified

b. [number: plural] = plural
[number:] = underspecified

c. **[gender: common] = common**
[gender:] = underspecified

d. **[animacy: human] = human**
[animacy:] = underspecified

e. *default* = indefinite, singular, neuter, non-human

I follow a proposal put forward by Schoorlemmer (2009) and take the representation of features in syntax to be entirely dependent on the morphological realization of these features: *morpho-driven feature specification*. Put differently, the lack of a value for an attribute (an underspecified feature representation) corresponds to a morphologically unrealized feature. For example, Dutch *singular* is morphologically unrealized and therefore mapped onto an attribute without a value. Dutch *plural* on the other hand is morphologically realized (by an *-s* or an *-en* morpheme) and thus mapped onto a fully specified syntactic feature representation, i.e. an attribute-value structure. This is illustrated in (63).[43]

(63)

interpretation	morpheme	feature structure
'singular'	–	[number:]
'plural'	*-s*/*-en*	[number: plural]

I follow Rooryck (2003) and Leu (2008) among many others and take definiteness to be morphologically realized by a *d*-morpheme.[44] Consequently, *definite*

[40] I take *human* (instead of the more broader term *animate*) to be the relevant specification for animacy here (cf. van Kampen 2010), but see section 2.5.2.6 for a more fine-grained classification of animacy.

[41] I do not think that *animacy* and *gender* can be reduced to one another (see e.g. Harley and Ritter 2002, and Toebosch 2007 for deficient pronouns in Dutch), because *neuter* as well as *common* gender nouns can be *animate* as well as *inanimate* in Dutch (cf. *infra*).

[42] Notice that this means that semantic features need not only be accessible at LF, but can be accessible at PF as well.

[43] This morpho-driven feature specification system fares better than an obligatory (under)specification system from a learnability point of view. Since features are only mapped onto an attribute when their interpretation is morphologically realized, the presence of a value for an attribute for which there is no morphological evidence does not need to be stipulated (see Schoorlemmer 2009 for details).

[44] Rooryck (2003:4) notes that – putting aside morpheme ordering between the *d*-morpheme and the schwa (ə) morpheme – the only difference between the definite determiners [d-ə] (*de*) 'the' (for common gender nouns) and [ə-t] (*'t*/*het*) 'the' (for neuter gender nouns) is that the *d*-morpheme in the latter has undergone final devoicing, which is characteristic of Dutch.

is syntactically represented by means of an attribute-value pair, whereas *indefinite* maps onto the attribute (underspecification). Supporting evidence for this way of encoding definiteness in the syntax is provided by bare nouns, which – in a system that has a morphological opposition between bare and non-bare nouns – are always indefinite (cf. Rooryck 2003). For example, in many languages – among which Dutch and English – indefiniteness in plural nouns is expressed by the bare plural noun, whereas definiteness in plural nouns is expressed by the plural noun in combination with the definite article (*boeken/books* vs. *de boeken/the books*). In sum, definiteness is more marked cross-linguistically.

As for the syntactic representation of gender in Dutch, I deviate from the proposal made in Schoorlemmer (2009), and instead follow a proposal by Kester (1996), Rooryck (2003), Toebosch (2007), van Kampen (2007) and BKL amongst others. I take neuter gender to be the underspecification for gender.[45,46] The reason for this is that neuter seems to function as the default value in Dutch: the neuter definite article *het* 'the' also functions as a pronominal expletive and personal pronoun, and the neuter demonstrative pronoun *dat* 'that' also seems to function as declarative complementizer (see chapter 4 for details). Moreover, when we focus on adjectival inflection, only adjectives in attributive position of neuter nouns do not get an overt agreement affix whenever the indefinite determiner *een* 'a/an' is used (cf. Kester 1996). This is illustrated in (64). Put differently, neuter is not morphologically realized, suggesting it is the default value for gender, and should thus map onto an attribute without a value.

(64) a. het mooi-**e** meisje — de mooi-**e** jongen
the.N beautiful girl.N — the.C beautiful boy.C

b. een mooi meisje — een mooi-**e** jongen
a beautiful girl.N — a beautiful boy.C

[45] Cf. Harley and Ritter (2002) who take neuter to be the default interpretation of CLASS (i.e. gender/animacy) *universally*.

[46] According to Schoorlemmer (2009:122) common gender is the underspecification for gender, i.e. common gender maps onto a gender attribute (not a value). The reason for this is that the definite article *de* 'the', which is used with common nouns, is also used with plural nouns, in which case *de* occurs with both common and neuter nouns. The article *de* is thus argued to not spell out common as it also occurs with neuter nouns in the plural. Instead, the article *het* 'the' is assumed to morphologically realize neuter. This is given in (i).

(i) [gender:] = common (underspecified, also compatible with neuter)
[gender: neuter] = neuter

Since gender is morphologically visible only on singular definite determiners and demonstratives, not on plural determiners (nor on the indefinite determiner), it is unclear which gender is morphologically realized: neuter or common. Moreover, as pointed out to me by Gertjan Postma, *within the CP-domain* – the domain in which events are encoded (and that is targeted by A-bar pronouns) – neuter is most likely indeed the default value, as *dat* is the element used to refer to events (*dat/*die heb ik niet gehoord* 'that.N/that.C' have I not heard).

As for *agreement* within this underspecification approach, I adopt the idea that agreement is feature sharing (cf. Pollard and Sag 1994, Frampton and Gutmann 2000, Pesetsky and Torrego 2007, Schoorlemmer 2009 a.o.). The following table illustrates gender agreement between a determiner (*probe*) and a noun (*goal*). Assuming that Dutch determiners are gender probes, they enter the derivation with an *underspecified* gender feature (cf. Chomsky 2001, Schoorlemmer 2009 a.o.). In accordance with the feature specification of gender in Dutch, cf. (62c), a *neuter* noun has an underspecified gender feature, whereas a *common* gender noun has a gender feature that is specified as *common*. As a result of agreement between the noun and the determiner, they share the gender feature that originated on the noun. Notice that agreement between two underspecified features is possible (i.e. agreement is feature sharing, independent of value): it does not cause the derivation to crash, and results in *default* morphology. The syntactic as well as the morphological outcome of this agreement relation is illustrated in the two rightmost columns of table 2.3. See section 2.5.3 for a detailed outline of how spell out of morphosyntactic features exactly works.

Table 2.3: Gender agreement between a determiner (D) and a noun (N)

D (probe)	N (goal)	syntactic outcome	morphological outcome
[gender:]	[gender:]	[gender:]	default morphology (= *neuter*) e.g. *het/dat meisje* 'the.N/that.N girl.N'
[gender:]	[gender: common]	[gender: common]	e.g. *de/die jongen* 'the.C/that.C boy.C'

Recall that in this chapter, I assume a traditional Head External Analysis (HEA) of RCs, according to which the RC head is base-generated in a position outside of the RC, and inside the RC a relative pronoun or operator moves to the left periphery where it is linked to the external head by means of predication (see chapter 3 for a specific implementation of a HEA of RCs). This is illustrated in (65).

(65) $\ldots$ RC head$_i$ [$_{RC}$ [REL.PRONOUN$_i$/OPERATOR$_i$]$_1$ $\ldots$ $\__1$ $\ldots$]

Instead of assuming that there is an agreement relation (in terms of feature sharing) between the RC head and the relative pronoun in RCs – notice that under a HEA of RCs as in (65) this would be a case of *Reverse Agree*: the RC head (goal) c-commands the relative pronoun (probe)[47] – I would like to propose that there is a feature matching requirement between the RC head and

[47] Zeijlstra (2010) proposes that agreement is always of this form, i.e. the probe (with [uF]) is c-commanded by the goal (with [iF]): [iF]>[uF](>[uF]).

the relative pronoun in the left periphery of the RC. The reason for this is that a HEA of RCs is hard to combine with an agreement proposal.

First, in long-distance RCs that show (identical) doubling, the form of the copy of the relative pronoun in the lower CP domain – indicated by underlining in (66) – is dependent on the features of the RC head.

(66) ... RC head $[_{CP}$ **RP**$_1$ ... $[_{CP}$ **<u>RP</u>**$_1$ ... ~~RP~~$_1$ ...]]

If there was an agreement relation between the RC head and the copy of the relative pronoun in the lower CP domain, this agreement relation would thus need to be able to feed into PF. If it did, the agreement approach to the relation between a relative pronoun and the RC head under a HEA would be incompatible with phase-based cyclic Spell-Out (cf. Chomsky 2001). That is to say, since the copy of the relative pronoun in the lower CP domain is in the complement of the higher phase head (C^0), under a phase-based cyclic Spell-Out approach it would have already been spelled out before the head of the movement chain of the relative pronoun agrees with the RC head. Consequently, feature sharing between the RC head and the copy of the relative pronoun in the lower CP domain should be invisible, contrary to fact.

Second, if there was an agreement relation between the RC head and the relative pronoun, it would be unclear how the ϕ-features on T^0 are valued in subject RCs. The RC head cannot do it because it is outside of the CP phase and T^0 is not in the edge domain of this CP phase. Under the assumption that agreement is subject to the *Phase Impenetrability Condition* (PIC, cf. Chomsky 2000 *et passim*; but see Bošković 2007), the PIC prevents a direct agree relation between the RC head and T^0.[48] The relative pronoun cannot value the ϕ-features on T^0 either, because it enters the derivation with underspecified ϕ-features itself, i.e. it is a probe for agreement (just like a determiner is a probe for gender agreement, as was illustrated in table 2.3).

So, to overcome both difficulties, I stipulate a feature matching requirement between the RC head and the pronoun in the left periphery of the RC. The relative pronoun thus does not act as a probe for agreement, as a result of which it can (i) be spelled out in the lower left periphery without having to wait for the merger of the RC head (i.e. it is compatible with phase-based cyclic Spell-Out) and (ii) value the ϕ-features on T^0 in case of a subject RC.

[48] See Heck and Cuartero (2008) for a tentative proposal to solve this particular problem. They argue that it is the RC head itself that values the ϕ-features on T^0, by assuming that agreement applies cyclically and involves feature sharing.

2.5.2.2 The feature specification of *wat*

As can be seen in (67) below (cf. table 2.2), the *w*-pronoun *wat* 'what' can have many different functions: it can be a quantifying determiner (67a), a relative pronoun (67b), an interrogative pronoun (67c), an exclamative marker (67d) and an indefinite pronoun (67e).

(67) a. Dat meisje heeft **wat** mensen geroepen.
that girl has what people called
'That girl called some people.'

b. Dat is het meisje **wat** die mensen heeft geroepen.
that is the girl what those people has called
'That is the girl who called those people.'

c. **Wat** heeft dat meisje gedaan?
what has that girl done
'What has that girl done?'

d. **Wat** een onzin!
what a nonsense
'What a nonsense!'

e. Dat meisje heeft **wat** geroepen.
that girl has what called
'That girl has called something.'

For *wat* to occur in so many environments, it must be highly *underspecified*. In fact, I take *wat* to be completely underspecified (cf. Postma 1994, Bennis 1995 and BKL a.o.), as illustrated in (68).

(68) feature specification of *wat*

a. [referentiality:] = underspecified for definiteness
b. [number:] = underspecified for number
c. [gender:] = underspecified for gender
d. [animacy:] = underspecified for animacy

First, *wat* is underspecified for definiteness: it may occur in an expletive construction without giving rise to a definiteness effect, as illustrated in (69).

(69) *Er* is **wat** gevonden.
there is what found
'Something was found.'

To illustrate further the correctness of the feature specification for *wat* in (68), consider table 2.4. This table illustrates for each of its uses, the antecedents that *wat* is compatible with (with respect to gender, number and animacy).[49] As this

[49] As an exclamative marker, *wat* is compatible with all antecedents (and can be accompanied by the spurious indefinite article *een* 'a/an'), cf. Bennis et al. (1998) and den Dikken (2006a) a.o. for details on *wat*-exclamatives and related constructions.

table shows, *wat* is compatible with both singular and plural antecedents (68b), with both neuter and common gender antecedents (68c), and with both human and non-human antecedents (68d). More specifically, as a relative pronoun, *wat* is compatible with singular neuter gender antecedents in Standard Dutch (70a), and in colloquial Dutch *wat* is compatible with singular non-human antecedents as well (70b). Whenever pronoun *wat* appears on its own (no RC antecedent), it gets the default *non-human* interpretation, i.e. as an indefinite or interrogative pronoun, *wat* can only refer to *non-human* referents.

(70) a. het meisje/boek **wat** hij mooi vond
the girl.N/book.N what he beautiful found
'the girl/book he found beautiful'

b. % de fout **wat** hij maakte
the mistake.C what he made
'the mistake he made'

Table 2.4: Combinatorial possibilities of *wat*

nominal antecedent/referent	determiner *wat*	relative pronoun *wat*	indefinite pronoun *wat*	interrogative pronoun *wat*
[N, SG, HUMAN] e.g. *meisje* 'girl'	−	+/%[50]	−	−
[N, SG, NON-HUMAN] e.g. *boek* 'book'	−	+/%	+	+
[C, SG, HUMAN] e.g. *man* 'man'	−	−[51]	−	−
[C, SG, NON-HUMAN] e.g. *fout* 'mistake'	−	%	+	+
[PL, HUMAN] e.g. *mannen* 'men'	+	−	−	−
[PL, NON-HUMAN] e.g. *boeken* 'books'	+	−	+	+
− = not attested, + = Standard Dutch, % = colloquial Dutch				

[50] According to the ANS (Haeseryn et al. 1997:339) the use of *wat* as a relative pronoun with neuter gender antecedents like *meisje* 'girl' and *boek* 'book' is *informal* Dutch. However, the MPQ1-A data show that almost all informants accept the use of *wat* as a relative pronoun for the neuter gender antecedent *boek* (417/452=92%). This suggests that we might be dealing with Standard Dutch after all.

[51] As mentioned in section 2.3, *wat* only occurs very marginally as a relative pronoun with common gender human antecedents like *man* 'man'.

The element *wat* in Dutch thus indeed seems to be completely underspecified.[52] It only contains an operator.

A small proviso is in order here. As becomes clear from table 2.4, the distribution of *wat* as a determiner is very different from the distribution of *wat* as a relative pronoun. Whereas determiner *wat* only occurs with plural (and mass noun) antecedents, relative pronoun *wat* only occurs with neuter and/or singular non-human antecedents. Put differently, there is no subset/superset relation between *wat* as a determiner and *wat* as a relative pronoun (cf. the patterns of the + and – signs in table 2.4). For this reason it is very complicated (if not impossible) to come up with a mechanism that accurately accounts for the interpretation and distribution of *wat* in all its different uses, given the feature specification in (68). Since I am only concerned with the use of *wat* as a pronoun, I leave the issue for further research of whether or not it is possible to come up with a mechanism that accounts for all the different uses of *wat*, while maintaining a single underspecified feature specification. For now, I therefore simply postulate that there is more than one lexical entry for *wat* after all, namely a lexical entry for determiner *wat* that is specified for plural/mass, and a lexical entry for pronoun *wat* that has the specification as given in (68) and that contains an operator. In the remainder of this chapter, I only deal with pronoun *wat*.

2.5.2.3 The feature specification of *wie*

Pronoun *wie* is more specified than pronoun *wat*, but it is still reasonably underspecified. As we saw in table 2.2, *wie* in Dutch is only used as an interrogative pronoun or a relative pronoun (in colloquial Dutch).[53] Being a *w*-pronoun, *wie* is only sensitive to *animacy* (van Kampen 2007, 2010). That is to say, *wie* is only compatible with human antecedents, irrespective of the gender or number of the antecedent, as can be seen in table 2.5. This is furthermore illustrated in (71) for relative pronoun *wie*. As an interrogative pronoun, *wie* can only question a person, i.e. the answer to a *wie*-question can only be a human referent.

(71) % het meisje/ de man/ de mannen **wie** ik gezien heb
the girl the man the men who I seen have

Therefore, I assume that *wie* has the following feature specification.

[52] This claim is further corroborated by the observation that in child language *wat* functions as the default relative pronoun for antecedents of which the gender has not yet been acquired (van Kampen 2007).

[53] I have no explanation for the observation that *wie* – unlike *wat*, cf. section 2.5.2.2 – does not allow for the indefinite reading ('someone'). See Postma (1994) for some speculation on this matter.

(72) feature specification of *wie*

a. [referentiality:] = underspecified for definiteness
b. [number:] = underspecified for number
c. [gender:] = underspecified for gender
d. [animacy: human] = human

Table 2.5: Combinatorial possibilities of *wie*

nominal antecedent/referent	relative pronoun *wie*	interrogative pronoun *wie*
[N, SG, HUMAN] e.g. *meisje* 'girl'	%	+
[N, SG, NON-HUMAN] e.g. *boek* 'book'	–	–
[C, SG, HUMAN] e.g. *man* 'man'	%	+
[C, SG, NON-HUMAN] e.g. *fout* 'mistake'	–[54]	–
[PL, HUMAN] e.g. *mannen* 'men'	%	+
[PL, NON-HUMAN] e.g. *boeken* 'books'	–	–
– = not attested, + = St. Dutch, % = colloquial Dutch		

2.5.2.4 The feature specification of *die*

Just like *wat*, *die* can be many things, and therefore must be fairly underspecified as well. The sentences in (73) show that *die* can be a demonstrative determiner (73a), a demonstrative (or resumptive) pronoun (73b), and a relative pronoun (73c).

(73) a. Ik heb **die** man niet gezien.
I have that.C man not seen
'I have not seen that man.'

b. (Jan,) ik heb **die** niet gezien.
Jan, I have that.C not seen
'(Jan,) I have not seen him.'

[54] See van der Horst and van der Horst (1999:172-173), Bennis (2001:19) and van der Horst (2008:1684) for some examples of relative pronoun *wie* referring to non-human antecedents, e.g. *de fout wie hun eigenlijk maken*, lit.: 'the mistake *who* they actually make' (Johan Cruijff, Studio Sport 1.5.2001; Bennis 2001:19). As most attestations of relative pronoun *wie* that refer to non-human antecedents come from the speech of a single person, namely Johan Cruijff, I disregard examples of *wie* referring to non-human antecedents in table 2.5.

c. Dat is de man **die** ik gezien heb.
that is the man RP I seen have
'That is the man who I have seen.'

As for its feature specification, consider table 2.6. This table shows that *die* seems to be the default pronoun in RCs, in the sense that it may occur with (almost) all antecedents.[55] A similar observation is made by Audring (2009:100): "the common gender demonstrative plays a central role in spoken Dutch. Especially the distal variant *die* is immensely frequent, and it can be used with nearly any noun". This is furthermore illustrated for *die* as a relative pronoun in (74).

Table 2.6: Combinatorial possibilities of *die*

nominal antecedent/referent	determiner *die*	relative pronoun *die*	demonstrative pronoun *die*
[N, SG, HUMAN] e.g. *meisje* 'girl'	–	%	%
[N, SG, NON-HUMAN] e.g. *boek* 'book'	–	%	%
[C, SG, HUMAN] e.g. *man* 'man'	+	+	+
[C, SG, NON-HUMAN] e.g. *fout* 'mistake'	+	+	+
[PL, HUMAN] e.g. *mannen* 'men'	+	+	+
[PL, NON-HUMAN] e.g. *boeken* 'books'	+	+	+
– = not attested, + = Standard Dutch, % = colloquial Dutch			

(74) a. de man/ de fout **die** ...
the man the mistake RP

b. de mannen/ de boeken **die** ...
the men the books RP

c. % het meisje/ het boek **die** ...
the girl the book RP

Instead of assuming that *die* is completely underspecified, on the basis of its distribution as a determiner – which neatly forms a subset of its distribution as

[55] Audring (2009:102) observes that the occurrence of a *neuter* gender mass noun (e.g. *brood* 'bread') followed by a *common* gender demonstrative pronoun (*die*) hardly ever occurs in Dutch. More specifically, common gender demonstrative pronouns *die* and *deze* cannot generally refer to neuter *unspecific* mass nouns or neuter *unbounded abstracts*. See section 2.5.2.6 for more details.

a pronoun – I assume that *die* is in fact specified as *common* and *definite*.[56,57] To illustrate why, consider the contrast in (75), cf. table 2.6.

(75) a. <dat/*die> meisje heb ik gezien
that.N/that.C girl.N have I seen

b. het meisje <dat/%die> ik gezien heb
the girl.N that.N/that.C I seen have

As a determiner, *die* cannot combine with a neuter noun, but as a relative pronoun it can; speakers differ in how good they consider (75b) to be, but all speakers have a clear contrast between (75a) and (75b).[58,59]

The *Agreement Hierarchy* as proposed by Corbett (1979, 1991, 2006) and here illustrated in (76), states that depending on the syntactic distance between the pronoun and the agreement controller, a pronoun is more likely to agree syntactically (i.e. agreement determined by the formal properties of the agreement controller) or semantically (i.e. agreement determined by the semantic properties of the agreement controller).

[56]Unlike what was the case with determiner *wat* and pronoun *wat* (cf. section 2.5.2.2), determiner *die* and pronoun *die* are in a subset/superset relation. This strongly indicates that there is a single lexical entry for *die*.

[57]It has been argued that *die* (and *d*-pronouns more generally) is not inherently definite, as it may appear in expletive constructions that otherwise trigger a strong definiteness effect. This is illustrated in (i)-(ii).

(i) de problemen **die** er zijn
the problems RP there are

(ii) wat/wie denk je **die** er zullen komen?
what/who think you RP there will come [den Dikken 2010:311]

Barbiers et al. (2010) show that *die* certainly is definite, and that (i) and (ii) are not foolproof tests to determine whether or not an element is definite. That is, they show that under the right circumstances, like *focus*, *modification* or *fronting*, *die* and the expletive *er* may co-occur. This is illustrated for fronting in (iii).

(iii) a. ***Die*** **problemen** zijn er nog steeds.
those problems where there yet still

b. Die problemen? ***Die*** zijn er nog steeds.
those problems those were there yet still

[adapted from Barbiers et al. 2010:313]

The lack of definiteness effect with *die*, as illustrated in (i) and (ii), might thus be reduced to the lack of definiteness effect in fronting constructions more generally. The claim that *d*-pronouns are *definite* can then be maintained.

[58]Expressions like *die meisje* 'that.C girl.N', as judged ungrammatical in (75a), are typically attested in the speech of second language (L2) learners of Dutch, but (almost) never in the speech of Dutch native speakers (L1).

[59]The observation that the same contrast is less clear when the neuter gender noun *boek* 'book' is involved has most likely to do with the fact that *meisje* 'girl' is *highly individuated*, whereas *boek* is not (see section 2.5.2.6 for details).

(76) **Agreement Hierarchy** (Corbett 1979, 1991, 2006)
attributive > predicate > relative pronoun > personal pronoun
syntactic *semantic*

Given that the syntactic distance (in terms of hierarchical distance) between the determiner and the noun *meisje* in (75a) is smaller than the syntactic distance between the relative pronoun and the noun *meisje* in (75b), the first is predicted to show syntactic agreement, whereas the latter does not have to. Assuming that agreement for gender (*neuter* or *common*) is syntactic in nature, we have an explanation for the pattern in (75) and the pattern in table 2.6 more generally. Determiner *die* occurs in attributive position and is thus only allowed to agree for gender with its agreement controller. Relative/demonstrative pronoun *die*, on the other hand, does not occur in attributive position, and therefore, does not have to agree for gender.[60]

The feature specification I assume for *die* is given in (77).

(77) feature specification of *die*
a. [referentiality: definite] = definite
b. [number:] = underspecified for number
c. [gender: common] = common
d. [animacy:] = underspecified for animacy

As mentioned in section 2.5.2.1, I do not assume there to be an *agreement* relation between the RC head and the relative pronoun in RCs. Rather the features on the relative pronoun need to *match* the features on the RC head. Therefore, I will not use the terms syntactic *agreement* and semantic *agreement*, but instead speak of *spelling out* syntactic features (gender) and *spelling out* semantic features (animacy). It should be mentioned here that spelling out semantic animacy in terms of humanness (human/non-human) is somewhat too simplistic. After the next subsection, which presents the feature specification of pronoun *dat* 'that', section 2.5.2.6 therefore briefly mentions a more complex system of semantic animacy that involves more features than only *humanness*.

2.5.2.5 The feature specification of *dat*

Just like pronoun *die*, pronoun *dat* can be a demonstrative determiner (78a), a demonstrative (or resumptive) pronoun (78b), and a relative pronoun (78c). In addition, *dat* is identical in form to the finite declarative complementizer, as illustrated in (78d). The discussion of the nature and status of complementizer *dat* and its relation with (relative) pronoun *dat* is postponed until chapter 4.

[60] Exactly the same holds for the element *dat*. As a determiner, *dat* may only occur with *neuter* gender antecedents, whereas its distribution is less restricted when it is used as a demonstrative or relative pronoun (cf. table 2.7 in section 2.5.2.5).

(78) a. Ik heb **dat** boek niet gelezen.
I have that.N book not read
'I have not read that book.'

b. (Dat boek,) ik heb **dat** niet gelezen.
that book, I have that.N not read
'(That book,) I have not read it.'

c. Dat is het boek **dat** ik heb gelezen.
that is the book RP I have read
'That is the book that I have read.'

d. Jan denkt **dat** ik het boek heb gelezen.
Jan believes that I the book have read
'Jan believes that I have read the book.'

In Standard Dutch, pronoun *dat* is only used with singular neuter gender antecedents, but in colloquial Dutch, pronoun *dat* can also be used with common gender or plural antecedents (cf. van der Horst 2008, Audring 2006, 2009 a.o.). This is illustrated in (79)-(80). Notice that the use of neuter pronouns with common gender antecedents is rather rare (Audring 2006), and that common gender nouns that are referred to by neuter pronouns are often *low individuated* (79) and/or *non-countable* (80) (cf. section 2.5.2.6).

(79) a. % Feyenoord is natuurlijk een ploeg **dat** zoiets aankan.
Feyenoord is of course a team.C that.N something like that handle
'Feyenoord is of course a team that can handle something like that.'

b. % de meeste bomen **dat** hier staan
the most trees.PL that.N here stand
'the most trees that stand here' [van der Horst 2008:1687]

(80) a. % dat er geen apparatuur onbeheerd is achtergebleven **dat** aan
that there no equipment.C unattended is remained that.N on
staat
stands
'that no equipment that is powered is left unattended'

b. % na het drinken van limonade **dat** met water was aangelengd
after the drinking of lemonade.C that.N with water was diluted
'after drinking the lemonade that was diluted with water'
[Audring 2006:80-81]

Table 2.7 illustrates all combinatorial possibilities of *dat* in Standard Dutch as well as in colloquial Dutch. As can be seen from this table, *dat* is underspecified for number, gender, and animacy. When it appears on its own it will get a *singular*, *neuter*, *non-human* interpretation by default. The feature specification of *dat* is given in (81).

(81) feature specification of *dat*

a. [referentiality: definite] = definite
b. [number:] = underspecified for number
c. [gender:] = underspecified for gender
d. [animacy:] = underspecified for animacy

Table 2.7: Combinatorial possibilities of *dat*

nominal antecedent/referent	determiner *dat*	relative pronoun *dat*	demonstrative pronoun *dat*
[N, SG, HUMAN] e.g. *meisje* 'girl'	+	+	+
[N, SG, NON-HUMAN] e.g. *boek* 'book'	+	+	+
[C, SG, HUMAN] e.g. *man* 'man'	–	_[61]	–
[C, SG, NON-HUMAN] e.g. *fout* 'mistake'	–	%	%
[PL, HUMAN] e.g. *mannen* 'men'	–	–	–
[PL, NON-HUMAN] e.g. *boeken* 'books'	–	%	%
– = not attested, + = Standard Dutch, % = colloquial Dutch			

2.5.2.6 The Individuation Hierarchy

Audring (2006, 2008, 2009) claims for Dutch that whenever a pronoun differs in syntactic gender (common/neuter) from its referent – i.e. there is a gender switch – the choice for the pronoun is based on what she refers to as *semantic gender*. That is to say, the form of the pronoun correlates with the degree of *individuation* of the antecedent: 'highly individuated' entities are best compatible with common gender pronouns, whereas 'low individuated' entities are best referred to with neuter pronouns. The *Individuation Hierarchy* as proposed in Audring (2009) (after Sasse 1993, Siemund 2008) is illustrated in (82).[62] Elements on the left of the hierarchy are highly individuated, whereas elements on the right of the hierarchy are low individuated. Below the hierarchy examples are given of each class of elements, and I furthermore roughly indicated

[61] There are varieties of Dutch in which subject RCs are introduced by *die* and object RCs are introduced by *dat*. In section 4.5, I show that in most of those dialects, *dat* is a finite declarative complementizer and not a relative pronoun. Whether or not *dat* can truly be an object specific relative pronoun in dialectal Dutch RCs is unclear, because it is very hard to distinguish complementizer *dat* from relative pronoun *dat* (see footnote 80 in chapter 4 for some discussion).

[62] The Individuation Hierarchy is a variant of the *Animacy Hierarchy* (Silverstein 1976).

the distribution of the relevant Dutch A-bar pronouns. As can be seen in (82), especially *die* has a wide distribution: it can be used with almost all *neuter* antecedents (in addition to being used with *common* gender and *plural* antecedents, cf. (74a)-(74b)).

(82) **Individuation Hierarchy** (Audring 2009:124)

male human/ female human	»	animal	»	bounded object/ abstract	»	specific mass	»	unspecific mass/ abstract
*man.*C ‘man’/ *meisje.*N ‘girl’		*kat.*C ‘cat’/ *paard.*N ‘horse’		*fout.*C ‘mistake’/ *boek.*N ‘book’		*thee.*C ‘tea’/ *brood.*N ‘bread’		*apparatuur.*C ‘equipment’/ *zand.*N ‘sand’
« —— »				« —— »				
wie				***die***		***dat***		***wat***

The Individuation Hierarchy may be formally implemented by means of the features *humanness* (human/non-human), *animacy* (animate/inanimate), *countability* (count/mass) and *boundedness* (bounded/unbounded).[63] As nothing really hinges on this for the discussion to come, I leave the formal implementation of the Individuation Hierarchy for future research. For now it suffices to say that in case syntactic gender (common/neuter) is not spelled out, semantic animacy *in accordance with the Individuation Hierarchy* as given in (82) can be spelled out. Concretely, this means that in case an A-bar pronoun is specified as [human], the lexical items *die* and *wie* are equally suited to spell out this A-bar pronoun.

[63] According to these four features, *wie*, *die*, *wat* and *dat* would be specified as follows (in a binary feature system):

wie =	[+human],	[+animate],	[+count],	[+bound]
die =	[+/-human],	[+/-animate],	[+/-count],	[+bound]
wat =	[-human],	[-animate],	[+/-count],	[+/-bound]
dat =	[-human],	[-animate],	[+/-count],	[+/-bound]

2.5.2.7 Interim summary and outlook

To summarize, *wat*, *wie*, *die* and *dat* have the following (simplified) feature specifications (features are between square brackets, underspecified features are omitted). I also indicated that all A-bar pronouns contain an operator.

(83) a. *wat*: operator
b. *wie*: operator, [human]
c. *die*: operator, [common], [definite]
d. *dat*: operator, [definite]

At this point it is important to note that *die* and *wie* do not *share* any features (although both of them contain an operator), so it cannot be the case that they are in a subset/superset relation, in contrast to what is claimed by BKL, cf. section 2.2.4. However, for many informants *die* and *wie* are in fact freely interchangeable in restrictive RCs with a human antecedent (and in the most deeply embedded clause of long-distance *wh*-Qs that question a person), suggesting that they do share some features. To account for their interchangeability in certain syntactic environments, I will argue that in a RC with a common gender human antecedent like *man* 'man' (as well as in the lower clause of a long-distance *wh*-Q that questions a person) *die* and *wie* fit the antecedent equally well. More specifically, both *die* and *wie* can equally well spell out the semantic feature [human]. In RCs, spelling out syntactic gender can thus be 'overruled' by spelling out semantic animacy.

The difference between spelling out syntactic gender (common/neuter) and spelling out semantic animacy (roughly human/non-human) in RCs is neatly reflected in the opposition between Standard Dutch and colloquial Dutch. Whereas the former requires spell out of syntactic gender, no such requirement holds in colloquial Dutch. This state of affairs is not unexpected: it makes sense that it is the more formal/normative language that requires the spell out of a grammatical distinction, and that it is the more informal language in which semantic notions like animacy can 'overrule' purely grammatical notions.

2.5.3 Lexicalization

With the feature specifications of the relevant A-bar pronouns in place, I now turn to the mechanism by which syntactic structures get spelled out by the right Vocabulary Items (lexicalization). The following table shows my assumptions for the spell out of a feature in an abstract manner. A feature representation [x:y] is best spelled out by a lexical item with the same specification: [x:y]. However, something that is specified in syntax ([x: y]) can in principle come out as underspecified in morphology ([x:]) as well, but only when there is no more specific form available (the more specific form *blocks* the less specific form, cf. the Elsewhere Condition (Kiparsky 1973)). The other way around, namely something underspecified in syntax ([x:]) comes out as specified in

morphology ([x: y]), I assume not to be possible. This entails that Vocabulary Items with underspecified feature representations have a wider distribution than Vocabulary Items with fully specified feature representations, as desired (cf. section 2.5.2 and chapter 4).

Table 2.8: Lexicalization

syntactic feature representation	Vocabulary Item	
[x: y]	[x: y]	= best match
[x: y]	[x:]	= possible match, unless blocked by the more specific [x: y]
[x:]	[x:]	= best match (default morphology)
[x:]	[x: y]	= no match

I propose to dissociate spell out of syntactic features and spell out of semantic features. More specifically, speakers of a language like Dutch, which exhibits grammatical gender distinctions within the A-bar pronoun system, may spell out grammatical gender (common/neuter), but may also choose to spell out animacy (roughy human/non-human), in accordance with the Individuation Hierarchy in (82).

First, let us abstract away from spelling out semantic animacy and only focus on spelling out syntactic gender. Consider therefore table 2.9, which shows lexicalization of syntactic gender in the domain of RCs. The leftmost column gives the feature specification of the 'relative pronoun' for gender (which matches the gender specification on the RC head, cf. section 2.5.2.1). The rightmost column illustrates if this feature representation can be spelled out by a lexical item (A-bar pronoun) that has the feature specification as given in the middle column. This table shows that common gender antecedents like *man* 'man' and *fout* 'mistake' will always be followed by the A-bar pronoun *die*, whereas a neuter gender antecedent will always be followed by the A-bar pronouns *dat* or *wat*.[64,65]

[64]The fact that the lexical item *wat* ([gender:]) is *in principle* compatible with the relevant syntactic feature specification ([gender: common]) – cf. table 2.8 – may account for the fact that *wat* occurs *very marginally* in RCs with a common gender human RC head (cf. the discussion following example (51c) in section 2.3).

[65]The observant reader may have noticed that the system of lexicalization as outlined in the main text predicts that *wie* – being underspecified for gender – can spell out the underspecified feature representation [gender:] as well. Whereas the fact that *wie* is inherently specified as [human] could be invoked to explain its non-occurrence with non-human antecedents like *boek* 'book' (**boek wie* 'book who'), it is unclear why in Standard Dutch *wie* cannot be the relative pronoun to a neuter gender human antecedent like *meisje* 'girl'. Perhaps *wie* is *un*specified for gender altogether, i.e. it completely lacks an attribute-value pair for gender (cf. Rooryck 1994 for the difference between *under*specification and *un*specification of features in syntax).

Table 2.9: Syntactic gender in RCs

syntax	lexicalization	
[gender: common]	[gender: common]	e.g. *man/fout die*
[gender: common]	[gender:]	e.g. **man/fout dat/wat/wie* (blocked by *die*)
[gender:]	[gender:]	e.g. *meisje/boek dat/wat*
[gender:]	[gender: common]	e.g. **meisje/boek die*

Now consider table 2.10, which shows lexicalization of semantic animacy in the domain of RCs. Like in table 2.9, the leftmost column gives the feature specification of the 'relative pronoun' for animacy (which matches the specification for animacy on the RC head, cf. section 2.5.2.1). The rightmost column illustrates if this feature representation can be spelled out by a lexical item (A-bar pronoun) that has the feature specification as given in the middle column. Recall from the Individuation Hierarchy in (82) that although *die* is not inherently specified as [human] – *die* also occurs with non-human antecedents in colloquial Dutch (e.g. *boek die* 'boek.N that.C') – it is compatible with a human antecedent (i.e. *die* and *wie* are equally compatible with a human antecedent). In RCs, a human antecedent like *man* can thus be followed by *wie* or *die*, whereas *dat* en *wat* are blocked in this environment. A non-human antecedent like *boek* on the other hand is compatible with *dat*, *wat* and *die*; only *wie* is excluded in such environments.

Table 2.10: Semantic animacy in RCs in line with Individuation Hierarchy

syntax	lexicalization	
[animacy: human]	[animacy: human]	e.g. *man/meisje wie/die*
[animacy: human]	[animacy:]	e.g. **man/meisje dat/wat* (blocked by *wie/die*)
[animacy:]	[animacy:]	e.g. *fout/boek dat/wat/die*
[animacy:]	[animacy: human]	e.g. **fout/boek wie*

Table 2.9 shows the pattern of relative pronoun selection as found in Standard Dutch, whereas table 2.10 shows all possibilities regarding relative pronoun selection that are attested in colloquial Dutch. Standard Dutch thus requires spell out of syntactic gender in RCs, whereas colloquial Dutch allows spell out of semantic animacy in RCs as well.

For ease of exposition, the following table summarizes the different outcomes of spelling out syntactic gender and spelling out semantic animacy. As mentioned before, treating the spell out of semantic animacy in terms of human versus non-human is not completely accurate. Spelling out semantic animacy must therefore always be taken to be spelling out semantic animacy *in accordance with the Individuation Hierarchy* as given in (82).

Table 2.11: Spell out of syntactic gender or semantic animacy in RCs

nominal antecedent (subject/object)	spell out of gender **Standard Dutch**	spell out of animacy (Individuation Hierarchy) **colloquial Dutch**
[animacy: human] [gender: common] e.g. *man* 'man'	**die**	**wie**/**die**
[animacy: human] [gender:] e.g. *meisje* 'girl'	**dat**/**wat**	**wie**/**die**
[animacy:] [gender: common] e.g. *fout* 'mistake'	**die**	**dat**/**wat**/**die**
[animacy:] [gender:] e.g. *boek* 'book'	**dat**/**wat**	**dat**/**wat**/**die**

We have now arrived at an explanation of the interchangeability of *wie* and *die* in RCs with a common gender human antecedent like *man* 'man' in colloquial Dutch. This is repeated in (84) below for doubling patterns in long-distance RCs.[66] Both *wie* and *die* match the [human] feature on the A-bar pronoun equally well (*wat* is blocked in this context).

(84) a. % Dat is de man [$_{RC}$ **die** ik denk [**die** het gedaan heeft]].
that is the man RP I think RP it done has
'That is the man who I think has done it.' = (51a)

b. % Dat is de man [$_{RC}$ **wie** ik denk [**wie** het gedaan heeft]].
that is the man who I think who it done has
'That is the man who I think has done it.' = (51b)

c. % Dat is de man [$_{RC}$ **wie** ik denk [**die** het gedaan heeft]].
that is the man who I think RP it done has
'That is the man who I think has done it.' = (52a)

d. % Dat is de man [$_{RC}$ **die** ik denk [**wie** het gedaan heeft]].
that is the man RP I think who it done has
'That is the man who I think has done it.' = (52b)

The same holds for the lower clause of a long-distance *wh*-Q in colloquial Dutch, as repeated here in (85).

[66] Taking doubling to be the repetition of a semantically superfluous element (cf. section 1.2.3), it is not surprising that doubling is primarily found in colloquial speech rather than in the standard language, as in the latter *normative pressure* to filter out semantically redundant material is higher (e.g. Weiß 2002, Alber 2008).

(85) a. % **Wie** denk je **wie** het gedaan heeft?
who think you who it done has = (19a)

b. % **Wie** denk je **die** het gedaan heeft?
who think you RP it done has = (19b)

More specifically, the A-bar pronoun to be lexicalized in a *wh*-Q that questions a person most likely contains an operator and the feature [human]. This A-bar pronoun can be realized by the lexical items *wie* and *die*. Although *wat* is in principle compatible with this structure as well – it matches the operator – it cannot be inserted into the lower clause of a long-distance *wh*-Q because it will lose the competition from *die* or *wie*, as these pronouns are better compatible with highly individuated entities like humans. This is in line with the facts: the grammaticality of (85a) and (85b) vs. the ungrammaticality of (86).[67,68]

(86) ?* **Wie** denk je **wat** het gedaan heeft?
who think you what it done has = (20a)

Following the null hypothesis that pronouns *die* and *wie* are the same elements in *wh*-Qs and in RCs, pattern *die-wie* in Qs (87) cannot be ruled out by the grammar as a violation of the *Inclusiveness Condition* (as argued by BKL, cf. (43a)), as this doubling pattern ís attested in RCs (88).

(87) * **Die** denk je **wie** het gedaan heeft?
RP think you who it done has = (20b)

(88) % Dat is de man [$_{RC}$ **die** ik denk [**wie** het gedaan heeft]].
that is the man RP I think who it done has
'That is the man who I think has done it.' = (52b)

The cause of the ungrammaticality of (87) – and the well known observation that *wh*-Qs cannot be introduced by *d*-pronouns more generally (cf. den Dikken and Bennis 2009) – I claim is a *wh*-requirement on the introduction of *wh*-Qs. More specifically, the C head introducing a *wh*-Q is endowed with a *wh*-feature that needs to be matched by the lexical item in its specifier. Assuming that *w*-pronouns, but not *d*-pronouns, bear a morphological *wh*-feature (with no interpretative relevance), only *w*-pronouns can satisfy that requirement. In the lower clause of long-distance *wh*-Qs on the other hand, no such requirement

[67]The observation that the highly underspecified *wat* is in principle compatible with a structure that contains an operator and the feature [human] – but loses the competition from *wie* and *die* – might account for the fact that doubling pattern *wie-wat* is not completely unattested in the Dutch speaking language area (cf. footnote 7).

[68]Notice that the sentence in (86) is also ruled out when the pronoun in the lower SpecCP is the spell out of an operator (cf. *infra*, section 2.5.4), and the pronoun in the higher SpecCP is the spell out of a DP that contains an operator and the feature [human]. More specifically, such a movement chain would violate the *Inclusiveness Condition* (Chomsky 1995:228), cf. BKL: features and structure are added to the operator in the course of the derivation.

exists and *wie* and *die* can equally well spell out the A-bar pronoun containing an operator and the feature [human].[69,70]

In sum, I provided an account of the doubling patterns involving *wie* and *die* in (long-distance) RCs with the common gender human antecedent *man*, and in *wh*-Qs that question a person. All doubling cases involve full movement of the A-bar pronoun and multiple copy spell out at PF. The different doubling patterns are simply the result of different lexicalizations of the feature bundle [operator, human, (common)]: *die* or *wie*. This is illustrated in (89) and (90).

(89) $[_{CP[+wh]}\ \underset{*die/\boldsymbol{wie}}{\mathrm{DP}_{[OP,HUMAN]}} \cdots\ [_{CP}\ \underset{\boldsymbol{die}/\boldsymbol{wie}/?*wat}{\mathrm{DP}_{[OP,HUMAN]}} \cdots$
long-distance (embedded) *wh*-question

(90) $[_{RC}\ \underset{\boldsymbol{die}/\boldsymbol{wie}/?*wat}{\mathrm{DP}_{[OP,HUMAN,C]}} \cdots\ [_{CP}\ \underset{\boldsymbol{die}/\boldsymbol{wie}/?*wat}{\mathrm{DP}_{[OP,HUMAN,C]}} \cdots$
long-distance restrictive relative clause

All doubling patterns in long-distance A-bar dependencies involving the pronouns *die* and/or *wie* thus have the exact same underlying syntax. The surface differences between them are determined at the level of morphology/PF.[71] How to account for the observation that *wat* may introduce *wh*-Qs that question a person is the topic of the next section.

2.5.4 Subextraction and spell out

Following a proposal by BKL, I assume that syntactic copying can be *partial*. That is to say, instead of copying a full constituent (*full* copying), the syntactic operation copying may also target a subconstituent and (re)merge it in a higher position (henceforth *subextraction*). In the structure of A-bar pronouns as proposed above, and repeated here as (91), this means that copying can either target the whole DP or a subpart of it, namely the operator in SpecDP.[72]

[69] This means that the *wh*-feature cannot be present in the feature bundle of the A-bar pronoun in syntax, as that would mean *die* could never occur in the lower clause of long-distance *wh*-Qs. Put differently, the *wh*-feature cannot be the driving force behind movement of the A-bar pronoun to the left periphery (*pace* den Dikken 2003 a.o.).

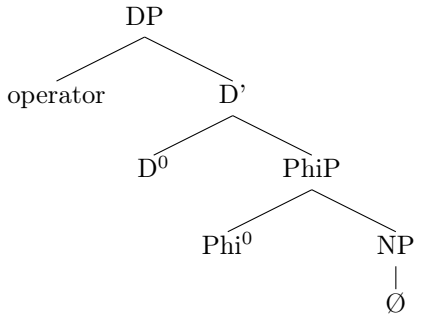
[70] Alternatively (or in addition), one may claim that the definiteness feature on *d*-pronouns is incompatible with a question interpretation, as a result of which *d*-pronouns cannot introduce questions. However, a problem with such an approach is that it immediately raises the question of how a Vocabulary Item that is inserted at PF could have semantic impact. More specifically, if the definiteness feature of an element comes in at PF, it is unclear how it could clash with the semantics of questions, given the model of grammar as briefly discussed in chapter 1.

[71] In the remainder of this chapter, I will refer to these doubling patterns (*wie-wie*, *die-die*, *wie-die*, and *die-wie*) as 'identical' doubling.

[72] Subextraction of the operator from the A-bar pronoun might be what underlies *wh-in-situ* (languages) – with the difference between overt *wh*-movement and *wh*-in-situ being that instead of spelling out the subextracted operator, only the A-bar pronoun itself is spelled out in-situ (and subextraction of the operator from thematic base position is possible).

Put differently, because the operator is the driving force behind movement to the left periphery, either it moves by itself (*subextraction*) or it *pied pipes* the entire DP.[73]

(91) The structure of A-bar pronouns (= (58))



The reason why both DP and the operator in SpecDP can be the target for copying, is that in their position in the lower SpecCP, they are *equally local* to the higher SpecCP (cf. *equidistance*, Chomsky 1995).[74,75] Recall that spell out of a copy in thematic base position is impossible (for whatever reason, cf. Nunes 2004, Thoms 2010 for some discussion), as a consequence of which subextraction from base position leads to a recoverability problem. Subextraction thus targets only elements in SpecCP.

At the point at which the operator inside the pronoun at the edge of the lower CP domain needs to move up, two possibilities emerge: either the whole pronoun (containing the operator that triggers movement) moves up, or only the operator itself moves up (subextraction). The two possible chains that we are left with are given in (92).

[73]In section 2.7.1, I will show that only the operator (spelled out as *wat* by default), but not larger phrases like *welke* 'which' can subextract – as illustrated in (i) – because subextraction of phrases like *welke* constitutes a violation of the Left Branch Condition (LBC, Ross 1967:207). The LBC states that extraction of a noun phrase on the left branch of another noun phrase leads to ungrammaticality. Notice that subextraction of the operator from SpecDP is not a violation of the LBC, because the operator is not a noun phrase. Even if subextraction of the operator were a violation of the LBC, it is ameliorated by spell out of the full DP from which it subextracted (see *infra*, section 2.5.5 for details).

(i) a. *Wat* denk je [[*welke man*] ik gisteren gezien heb]?
what think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'

b. * *Welke* denk je [[*welke man*] ik gisteren gezien heb]?
which think you which man I yesterday seen have

[74]Where *equally local* is formulated as follows: Y and Z are equally local to X if and only if (i) X c-commands both Y and Z, and (ii) the set of nodes that c-command Y is identical to the set of nodes that c-command Z (van Koppen 2005:14).

[75]This *subextraction* operation is clearly more restricted than the *partial copying* operation of BKL: it can only target the operator in SpecDP.

(92) a. $[_{CP}$ pronoun$_1$ ... $[_{CP}$ pronoun$_1$ ... pronoun$_1$ ...]]
full copying

b. $[_{CP}$ operator$_1$ ... $[_{CP}$ pronoun$_1$ ... pronoun$_1$ ...]]
subextraction

As we saw in section 2.2.3, the chain in (92a) can result either in the spell out of the highest copy (*no doubling*), or in the spell out of the highest *and* the intermediate copy (as the result of *morphological reanalysis* of the intermediate copy in SpecCP and the intermediate C head, cf. Nunes 2004). This is illustrated in (93).

(93) a. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ ~~pronoun$_1$~~ ... ~~pronoun$_1$~~ ...]]
no doubling

b. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ $[_{C}$ **pronoun**$_1$ $[_{C}$ Ø]] ... ~~pronoun$_1$~~ ...]]
doubling

As for the linearization or spell out of the chain in (92b), I assume that the operator becomes PF visible when extracted (cf. BKL and references cited therein). More specifically, when the operator is subextracted from the pronoun DP, it is spelled out as *wat*, because *wat* is the most underspecified A-bar pronoun in Dutch: it only contains an operator (cf. *supra*). Assuming that a single lexical item *wat* may spell out a full DP as well as an operator, suggests that lexicalization is governed by some sort of *Superset Principle*. The Superset Principle, as formalized in the Nanosyntax framework by Starke (unpublished work) and Caha (2007, 2009), is given in (94).[76]

(94) *The Superset Principle* (Caha 2007:3, cf. Ramchand 2008a,b)
The phonological exponent of a Vocabulary Item is inserted into a node if the item matches all or a superset of the grammatical features specified in the node. Insertion does not take place if the Vocabulary Item does not contain all features present in the node. Where several Vocabulary Items meet the conditions for insertion, the item containing less features unspecified in the terminal morpheme must be chosen.

According to the Superset Principle, a Vocabulary Item can thus spell out a syntactic structure that is smaller than itself. Put differently, the formulation of the Superset Principle in (94) entails that all syntactic features and syntactic structure should be lexicalized (only features of the lexical entries might be ignored). This is called *Exhaustive Lexicalization* (cf. Fábregas 2007 for discussion). Whenever a feature of a lexical entry does not match a feature in the syntactic structure, this feature is referred to as being *underassociated*. The Superset Principle thus gives us a handle for understanding the observation

[76] The Nanosyntax approach to language was initiated by Michal Starke and further developed at the University of Tromsø, cf. the collection of papers in Svenonius et al. (2009), and see `http://nanosyntax.auf.net/blog/`.

that one and the same pronoun can occur in different syntactic environments: features in the lexical entry of the pronoun can be underassociated in certain contexts, but the lexical entry of the pronoun itself is invariant, so there is no need to postulate multiple lexical entries for a single pronoun.

The Superset Principle is inherently incompatible with an *underspecification* approach to syntactic features. Elements that can have more than one function and may appear in more than one context, cannot be *underspecified*, but need to be *overspecified*. For example, being able to occur with singular as well as with plural antecedents, does not mean being underspecified for number, but rather being specified as [singular] *and* as [plural] (i.e. *overspecification*). We thus need an alternative mechanism of Vocabulary Item insertion that selects the Vocabulary Item that matches the *most features* in the feature bundle to be lexicalized (cf. the *Subset Principle*, standardly assumed in the Distributed Morphology framework, see footnote 77), while at the same time allows features of the Vocabulary Item to not match features in the syntactic structure (cf. *underassocation* and the Superset Principle). Such a principle should look something like (95).[77,78]

(95) *The Closest Match Principle*
The phonological exponent of a Vocabulary Item is inserted into a node if the item matches *one or more* of the grammatical features specified in the node. Where several Vocabulary Items meet the conditions for insertion, the item that matches the *greatest number* of features specified in the node and that contains the *smallest number* of features unspecified in the node must be chosen.

Whereas all A-bar pronouns contain an operator, as a result of which all of them are possible lexicalizations of the operator, the Closest Match Principle

[77]Notice that (95) basically is the Subset Principle (i) *minus* the condition that *insertion does not take place if the Vocabulary item contains features not present in the morpheme* and *plus* the idea that non-terminal nodes can be lexicalized as well (cf. Caha 2007 for discussion) – recall that I assume A-bar pronouns to spell out non-terminals (cf. section 2.5.1).

(i) *The Subset Principle* (Halle 1997)
The phonological exponent of a Vocabulary Item is inserted into a morpheme in the terminal string if the item matches all or a subset of the grammatical features specified in the terminal morpheme. Insertion does not take place if the Vocabulary Item contains features not present in the morpheme. Where several Vocabulary Items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen.

[78]If something like the Closest Match Principle is adopted throughout, the lexicalization of a single feature structure as sketched in the previous section (cf. table 2.8) changes a bit, in the sense that a feature that is underspecified or absent in syntax may *in principle* come out as specified or present in morphology (cf. *underassociation*), as long as there is no better matching element that would block this option. Since such a scenario never arises in the case of spelling out syntactic gender or semantic animacy features (because there is always a better matching element that blocks the insertion of an overspecified element), I leave this issue here. Needless to say, further research is required to determine the exact principles and constraints that underly lexicalization.

will select *wat* as the most optimal realization of the operator, because *wat* has the least features underassociated in its lexical entry (*least junk*). Put differently, the Closest Match Principle chooses *wat* as the best match for the operator.[79]

In addition to the operator in the higher SpecCP, the copy of the pronoun DP in the embedded SpecCP needs to be spelled out as well for reasons of *recoverability*, i.e. the features present in the intermediate copy need to be spelled out.[80] As spell out of the pronoun DP *subsumes* spell out of the operator – recall that pronouns are assumed to spell out phrases, cf. section 2.5.1 – this intermediate copy will always surface as a full pronoun. This is illustrated in (96).

(96) $[_{CP}$ **operator**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
subextraction

We thus have an account for the grammatical doubling patterns involving *wat* in a long-distance *wh*-Q that questions a person. We start out with a DP structure containing an operator and the feature [human]. This DP moves up to the lower SpecCP, from which the operator subextracts and moves to the higher SpecCP. The operator higher up is spelled out as *wat*, and the pronoun DP in the lower CP domain is spelled out as *wie* or *die*: both lexical items match the [human] feature equally well.[81] This is illustrated in (97) and exemplified by the sentences in (98).

(97) $[_{CP[+wh]}$ **operator** / ***wat*** ... $[_{CP}$ **DP**$_{[OP,HUMAN]}$ ... / ***wie*** / ***die***
long-distance (embedded) *wh*-question

[79]The fact that *d*-pronouns are inherently specified for definiteness, yet can be inserted in environments that are most likely not definite (i.e. RCs and the lower clause of *wh*-Qs), might be accounted for by the Closest Match Principle as well. That is to say, it might be exactly the *underassocation* of the definiteness feature of *d*-pronouns in RCs that ameliorates the definiteness effect in RCs. Put differently, the underassociation of the definiteness feature of *d*-pronouns in RCs might give an account of the well known lack of definiteness effect or 'the indefiniteness of the trace' in RCs (cf. Bianchi 1999, Bhatt 2002, Salzmann 2006, and Cinque 2008 a.o.).

[80]Notice that this 'multiple' spell out is not in violation of the LCA, as the copies are not (featurally) identical (operator vs. DP).

[81]A natural question that arises at this point is the following: if *die* is not inherently specified as human, then do sentences like (98b) actually always get a *human* interpretation? I am not convinced that this is the case. Rather, I think that whether or not the *wat-die* pattern gets a human interpretation is highly dependent on the type of verb involved. It might very well be the case that a sentence like (i) can be answered with *die acteur* 'that actor', as well as with *die film* 'that movie'. Future research should settle this issue.

(i) Ze vroeg **wat** jij denkt **die** een Oscar gewonnen heeft.
she asked what you think RP an Oscar won has
'She asked who/what you think won an Oscar.'

Notice that even if it turns out that sentences like (i) always get a human interpretation, one can build on the observation that *die* is most easily compatible with highly individuated entities like humans (cf. section 2.5.2.4).

(98) a. % **Wat** denk je **wie** het gedaan heeft?
what think you who it done has = (19c)

b. % **Wat** denk je **die** het gedaan heeft?
what think you RP it done has
'Who do you think has done it?' = (19d)

The ungrammaticality of the patterns in (99) was already accounted for in section 2.5.3. Either these sentences involve full copying and double spell out, in which case their ungrammaticality is explained by the fact that *wat* cannot spell out a DP that contains an operator and the feature [human], or these sentences are a violation of the *Inclusiveness Condition*: the feature [human] is added to the operator in the course of the derivation. Notice furthermore that (99b) is ruled out by the fact that *die* cannot satisfy the *wh*-requirement on the introduction of *wh*-Qs.

(99) a. ?* **Wie** denk je **wat** het gedaan heeft?
who think you what it done has = (20a)

b. * **Die** denk je **wat** het gedaan heeft?
RP think you what it done has = (20c)

At this point one might object that subextraction of the operator from DP violates well established constraints on movement. That is, subextraction as in (96) constitutes a violation of the *Condition on Extraction Domain* (CED, Huang 1982) or the *Freezing Principle* (Wexler and Culicover 1980, cf. also Corver 2006), according to which a phrase that has undergone movement becomes an island for extraction.[82,83] In order to obviate such locality violations one might delete the lower copy (*rescue by PF deletion*, cf. Bošković 2011 and see section 2.5.5). However, since deletion of the lower copy would lead to a recoverability problem, the lower copy needs to be spelled out, and in doing so, a violation of the CED/Freezing Principle is circumvented. Put differently, the pronoun in the lower SpecCP acts as an *intrusive* resumptive pronoun (Sells 1984) in the sense that it obviates a CED/Freezing Principle violation. I will

[82]The *Condition on Extraction Domain* (Huang 1982) is a condition of bounding theory – a theory about the locality of movement (cf. Chomsky 1981) – that states that extraction out of a domain D is possible only if D is properly governed. The CED basically states the impossibility of extraction out of certain constituents, like adjuncts (*Adjunct Condition*) and subjects (*Subject Condition*).

[83]As mentioned by Gallego (2010) (and see also Gallego and Uriagereka 2007), in more recent analyses of CED effects, subextraction from SpecCP is not necessarily problematic. Gallego (2010) argues that freezing effects have nothing to do with the structural configuration (phase edges), but rather with the position in which features are checked off. More specifically, as soon as an element reaches a position where its features are (fully) checked off, it is rendered opaque and hence cannot be targeted by subextraction. Since in the long-distance A-bar dependencies in the main text, movement to the embedded SpecCP does not check off features of the A-bar pronoun (rather it is forced by the uF feature on the operator/pronoun), under Gallego's (2010) analysis, subextraction from the A-bar pronoun may occur.

call this mechanism *rescue by PF spell out* – the logical counterpart of *rescue by PF deletion*. See section 2.5.5 for details.

In sum, I claim that doubling involving an instance of *wat* in the higher clause of a long-distance *wh*-Q that questions a person is the result of subextraction of the operator from DP. This operator is spelled out as *wat* and the DP (in the embedded SpecCP) from which it is extracted is spelled out as *die* or *wie* for reasons of recoverability and in order to obviate a violation of the CED/Freezing Principle. At this point it is unclear whether or not subextraction of the operator and double spell out is possible in RCs as well, because the doubling patterns in long-distance RCs with the common gender antecedent *man* 'man' do not feature the element *wat*. See section 2.6 for discussion.

2.5.5 Rescue by PF spell out

Ross (1969) was the first to argue that ellipsis may ameliorate island effects, as illustrated in (100) for sluicing. The example in (100a) shows that movement of *which one of my friends* violates the Complex NP Constraint (CNPC), giving rise to ungrammaticality. The example in (100b), on the other hand, shows that in case the category containing the island violation is deleted under ellipsis, the sentence becomes fine.

(100) a. * She kissed a man who bit one of my friends, but Tom does not realize which one of my friends she kissed a man who bit.

b. ? She kissed a man who bit one of my friends, but Tom does not realize which one of my friends.

[Ross 1969:276, cited in Bošković 2011:2]

Bošković (2011) proposes to extend the application domain of the *rescue by PF deletion* approach to all kinds of locality of movement violations. Locality of movement thus needs to be partly representational, as locality violations may be ameliorated at PF. Put differently, at least some aspects of locality of movement need to be attributed to PF (cf. also Pesetsky 1998 a.o. for a PF theory of locality). Most importantly for present purposes, Bošković argues that next to ellipsis, copy deletion may ameliorate island violations as well; this accounts for Chomsky's (1995, 2001) generalization that traces/unpronounced copies do not count as interveners for relativized minimality effects. To illustrate this claim, consider the sentences in (101), which show experiencer blocking in Italian. Whereas sentence (101a) shows that movement of *Gianni* across *a Maria* yields a relativized minimality violation (both are A-specifiers), sentence (101b) shows that when the copy inducing the violation is deleted, the sentence becomes grammatical. Island violations are indicated by a *; if a * remains in the final structure, the sentence is ungrammatical.[84]

[84]Bošković (2011:8ff.) assumes that a * is only copied under movement when the element that got assigned the * undergoes the same type of movement that has caused the violation (only relevant for relativized minimality violations). In (101b) the movement that causes the

(101) a. * Gianni$_1$ sembra a Maria* [$\__1$ essere stanco].
Gianni seems to Maria to be ill

b. A Maria$_2$, Gianni$_1$ sembra ~~a Maria$_2$*~~ [$\__1$ essere stanco].
to Maria Gianni seems to be ill
'To Maria, Gianni seems to be ill.' [Italian, Bošković 2011:4]

As mentioned in the previous section, I propose an operation that ameliorates movement violations at PF by means of spell out: *rescue by PF spell out*. More specifically, I take non-identical doubling involving *wat* 'what' in *wh*-Qs to be the result of subextracting the operator from the pronoun in the embedded CP domain. This subextraction is a violation of the CED/Freezing Principle. Since *deletion* of the offending copy (*rescue by PF deletion*) is not an option because it would give rise to a recoverability problem (deletion upon recoverability), I suggest that the copy is *spelled out* instead. By spelling out the full pronoun (containing a copy of the subextracted operator), a violation of the CED/Freezing Principle is *repaired*. Doubling in constructions in which the operator subextracts from an A-bar pronoun are thus predicted to be obligatory. This prediction is borne out, as illustrated in (102).

(102) Wat$_1$ denk je *(wie$_1$) het gedaan heeft?
what think you who it done has
'Who do you think has done it?'

A similar proposal has been made by van Craenenbroeck and van Koppen (2008) for first conjunct clitic doubling (FCCD) in southern Dutch dialects. In a FCCD sentence, the first conjunct of a coordinated subject is doubled by a clitic, as illustrated in (103).

(103) omda-ge gou en ik makannern gezien emmen
because-you$_{CLITIC}$ you$_{STRONG}$ and I each other seen have
'because you and I saw each other' [Wambeek Dutch]

Van Craenenbroeck and van Koppen (2008) analyze FCCD as follows. Part of the DP in the first conjunct of the coordination subextracts (namely PhiP) and is spelled out higher up as a clitic (*ge*), whereas the DP left behind by movement is spelled out as a strong pronoun (*gou*). However, subextraction from the first conjunct of a coordination violates the *Coordinate Structure Constraint* (CSC; Ross 1967:161), according to which extraction (out) of a conjunct in a coordinate structure is impossible. Van Craenenbroeck and van Koppen (2008) assume that this violation can be salvaged by spelling out the pronoun in the first conjunct. Doubling is then predicted to be obligatory, because the pronoun in the first conjunct acts as an intrusive resumptive pronoun that obviates the CSC violation. This prediction is borne out, as illustrated in (104).

violation is A-movement, whereas *a Maria* undergoes A-bar movement. The * is thus *not* copied onto *a Maria* under movement. So, after the * on the lower copy of *a Maria* is deleted under copy deletion, no *-marked element survives the deletion, and (101b) is grammatical.

(104) omda-ge *(gou) en ik makannern gezien emmen
because-you$_{CLITIC}$ you$_{STRONG}$ and I each other seen have
'because you and I saw each other' [Wambeek Dutch]

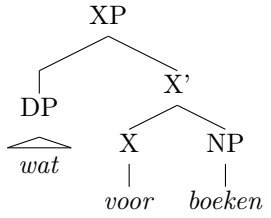
Additional evidence in favor of the claim that spell out may rescue an otherwise illicit step in the derivation (i.e. a violation of the CED/Freezing Principle), comes from the famous *wat voor* construction in Dutch. In case of doubling, this construction – just like pronominal doubling involving *wat* – also seems to involve subextraction from SpecCP.[85] The pattern of *wat voor* split in (105) shows that subextraction from the embedded SpecCP position is only licensed when the whole copy in the embedded SpecCP position is spelled out: (105b) versus (105c). Put differently, the full XP in (106) is spelled out after subextraction of the operator.[86] This seems to provide additional evidence in favor of *rescue by PF spell out*.

(105) a. [**Wat voor boeken**] denk je [dat hij heeft gelezen]?
what for books think you that he has read
'What kind of books do you think that he read?'

b. % **Wat** denk je [[**wat voor boeken**] hij heeft gelezen]?
what think you what for books he has read
'What kind of books do you think that he read?'

c. * **Wat** denk je [[**voor boeken**] hij heeft gelezen]?
what think you for books he has read

(106)



Just like with doubling involving A-bar pronouns, subextraction of the operator is not possible from base position (recall that *wh*-phrases cannot be spelled out in base position). If it were, double spell out (of the operator in the highest SpecCP and the A-bar pronoun in base position) would be required for recoverability reasons, and the construction in (107) would be grammatical, contrary to fact.

(107) * **Wat** denk je [dat hij [**wat voor boeken**] heeft gelezen]?
what think you that he what for books has read

[85] It is beyond the scope of this thesis to engage in an analysis of the *wat voor* (split) construction besides assuming it involves *subextraction* (but see Bennis 1983, 1995, Corver 1991, Bennis et al. 1998, den Dikken 2006a, Leu 2008 a.o.).

[86] For ease of exposition, I abstract away from the presence of the spurious indefinite article *een* 'a/an' in the *wat voor* construction (in the structure in (106)).

Similarly, subextraction of the operator (and double spell out) is impossible from the subject in SpecTP. This is illustrated by the sentences in (108): subextraction of the operator from the subject in SpecTP leads to ungrammaticality (108a), whereas subextraction of the operator from the subject in SpecCP is attested (108b). More generally, subextraction of the operator (and subsequent double spell out) is only possible from SpecCP, i.e. subextraction of the operator is only possible from an A-bar position. At this point, I have no insight to offer as to why this is the case.

(108) a. * **Wat** denk je [dat [**wat voor jongens**] dit boek hebben gelezen]?
what think you that what for boys this book have read

b. % **Wat** denk je [[**wat voor jongens**] dit boek hebben gelezen]?
what think you what for boys this book have read
'What kind of boys do you think read this book?'

I assume that PF spell out is only licensed in case PF deletion would lead to a recoverability problem. Interestingly, it is possible to subextract the A-bar pronoun *wat* from the *wat voor XP* in its thematic base position.[87] Since this subextraction does not lead to a recoverability problem (i.e. recoverability is ensured by spelling out the A-bar pronoun *wat* in the higher SpecCP and spelling out *voor XP* in base position), we thus predict that subextraction of the A-bar pronoun from base position does not lead to spell out of the full constituent from which subextraction takes place. This prediction is borne out, as illustrated by (109).[88] Notice that subextraction from base position does not constitute a violation of the CED/Freezing Principle, as a result of which (109) is perfectly grammatical.

(109) **Wat** denk je [dat hij [**voor boeken**] heeft gelezen]?
what think you that he for books has read
'What kind of books do you think that he read?'

[87]It is well known that the *wat voor XP* construction (as in (106)) allows subextraction of its specifier (*wat*), in violation of the Left Branch Condition (cf. Bennis 1983, 1995, den Besten 1985, Corver 1991, Corver 2003 a.o.). I have no insight to offer as to why this is the case.

[88]As pointed out to me by Hans Bennis, some speakers accept the sentence in (i). This is expected by my analysis of doubling: after subextraction from the complex *wh*-phrase in thematic base position, the A-bar pronoun successive-cyclically moves up to the higher SpecCP. If only the head of the movement chain of the A-bar pronoun in this construction is spelled out (i.e. the copy of the A-bar pronoun in the highest SpecCP), we get the construction in (109). However, if multiple copies of this A-bar pronoun movement chain are spelled out (i.e. the copy of the A-bar pronoun in the higher SpecCP and the copy of the A-bar pronoun in the lower SpecCP), we get the construction in (i). Alternatively, the construction in (i) can be derived by first subextracting the A-bar pronoun from the *wat voor XP* from thematic base position, and then subextracting the operator from the A-bar pronoun in the embedded SpecCP.

(i) % **Wat** denk je [**wat** hij **voor boeken** heeft gelezen]?
what think you what he for books has read
'What kind of books do you think that he read?'

Unlike subextraction of pronoun *wat* from direct object base position, subextraction of pronoun *wat* from subject position (SpecTP) is severely degraded (or ungrammatical), as illustrated in (110) (cf. Bennis 1995:32).

(110) ?* **Wat** denk je [dat [**voor jongens**] dit boek hebben gelezen]?
what think you that for boys this book have read
'What kind of boys do you think have read this book?'

The degraded grammaticality (or ungrammaticality) of (110) is in fact predicted: because subextraction of the A-bar pronoun from the *wat voor XP* in SpecTP does not lead to a recoverability problem (cf. *supra*), PF spell out cannot apply to salvage the CED/Freezing Principle violation that is caused by subextraction from derived position. The difference in grammaticality between subextraction of the A-bar pronoun from the direct object base position (109) and subextraction of the A-bar pronoun from the derived subject position (110) is thus explained in terms of the presence or absence of a CED/Freezing Principle violation. Similarly, the observation that it is impossible to subextract the A-bar pronoun from the *wat voor XP* in SpecCP, as illustrated in (105c), can be explained as follows: it constitutes a violation of the CED/Freezing Principle that cannot be overcome by PF spell out, because recoverability is ensured by spelling out A-bar pronoun *wat* higher up and *voor XP* lower down.

So, I take PF deletion to be more economical than PF spell out (cf. Nunes 2004 a.o.): only when PF deletion cannot apply due to the lack of recoverability (deletion upon recoverability), PF spell out can apply. This means that rescue by PF spell out can only salvage a derivation that involves *subextraction*, because only in that case can PF deletion lead to a recoverability problem (but this is not necessary, cf. *supra*). The only way to repair a derivation in such cases is to spell out the phrase from which an element has subextracted.

2.5.6 Summary

I have argued that doubling in long-distance *wh*-Qs and RCs is the result of either full movement and multiple copy spell out ('identical' doubling) or of subextraction and double spell out (non-identical doubling). More specifically, I have provided an account of the attested *grammatical* and *ungrammatical* doubling patterns in long-distance *wh*-Qs and in long-distance RCs with a common gender human antecedent. We have thus arrived at a theory of doubling patterns that *can* and doubling patterns that *cannot* be generated by the grammar. I leave for future (sociolinguistic) research the question of *why* a certain informant *realizes* one doubling pattern but not another (cf. section 1.2.2: *ungrammatical* vs. *unrealized* structures). Table 2.12 summarizes the analysis of the attested and unattested doubling patterns.

Table 2.12: Analysis of pronoun doubling in long RCs and *wh*-Qs in Dutch

pattern	restrictive relative clauses with antecedent *man* 'man' (common gender)	embedded *wh*-questions that question a person (*wie* 'who')
wie-dat	DP movement + deletion of all but highest copy	
wie-wie	DP movement + multiple copy spell out	
wie-die	DP movement + multiple copy spell out	
die-dat	DP movement + deletion of all but highest copy	independently ruled out: *die* cannot introduce *wh*-Qs
die-die	DP movement + multiple copy spell out	independently ruled out: *die* cannot introduce *wh*-Qs
die-wie	DP movement + multiple copy spell out	independently ruled out: *die* cannot introduce *wh*-Qs
wat-wie	independently ruled out: *wat* cannot be RP to *man*	subextraction + double spell out
wat-die	independently ruled out: *wat* cannot be RP to *man*	subextraction + double spell out
wie-wat	independently ruled out: *wat* cannot be RP to *man* or ruled out by *Inclusiveness*	independently ruled out: *wat* is blocked in lower clause or ruled out by *Inclusiveness*
die-wat	independently ruled out: *wat* cannot be RP to *man* or ruled out by *Inclusiveness*	independently ruled out: *die* cannot introduce *wh*-Qs, *wat* is blocked in lower clause or ruled out by *Inclusiveness*

2.6 Predictions and empirical support

The analysis of doubling in long A-bar dependencies presented in this chapter makes a number of clear predictions. This section discusses the predictions the analysis makes and provides empirical support for them.

2.6.1 Doubling in RCs with a neuter gender RC head

The proposed analysis makes clear predictions about the possible existence of certain doubling patterns. I argued that identical doubling with *wat* 'what' in long-distance RCs is not attested with the common gender human RC head *man* 'man', because *wat* cannot be a relative pronoun to a common gender human RC head in the first place. Doubling with *wat* is predicted to occur in long-distance RCs with a RC head that can combine with relative pronoun *wat*, namely a neuter gender RC head (cf. section 2.5.3). This prediction is borne out by the MPQ2-A data, as illustrated in (111) for long-distance subject RCs

with the neuter gender human RC head *meisje* 'girl' and the neuter gender non-human RC head *boek* 'book' respectively.

(111) a. % Dat is het meisje **wat** ik denk **wat** het gedaan heeft.
that is the girl what I think what it done has
'That is the girl who I think has done it.' [62/255=24%]

b. % Dat is het boek **wat** ik denk **wat** de prijs gewonnen heeft.
that is the book what I think what the prize won has
'That is the book that I think won the prize.'
[100/256=39%, MPQ2-A data]

Nothing in principle prohibits subextraction of the operator and subsequent double spell out that is found in *wh*-Qs to extend to RCs. The non-identical doubling patterns involving *wat* in long-distance *wh*-Qs should thus be replicated in long-distance RCs, under the condition that the operator in the higher SpecCP domain – spelled out by pronoun *wat* – matches the features on the RC head (cf. also section 2.5.2.1), i.e. *wat* can be a relative pronoun to the given RC head. This is suggested by the ungrammaticality of (112), under the assumption that these sentences are the result of subextraction of the operator and double spell out.

(112) a. ?* Dat is de man **wat** ik denk **wie** het gedaan heeft.
that is the man what I think who it done has = (54a)

b. ?* Dat is de man **wat** ik denk **die** het gedaan heeft.
that is the man what I think RP it done has = (54c)

More specifically, a neuter gender RC head that can occur with the relative pronoun *wat* as well as with the relative pronoun *die*, is predicted to show the doubling pattern *wat-die*, as this pattern would be the manifestation of subextraction of the operator (which is spelled out as *wat* by default) and double spell out. Crucially, the reverse pattern (*die-wat*) is predicted to be nonexistent as it would constitute a violation of the *Inclusiveness Condition*. This prediction is largely borne out, as illustrated in (113) and (114) for the neuter gender human RC head *meisje* 'girl' and the neuter gender non-human RC head *boek* 'book' respectively. Although the number of attestations of the a-sentences is not very high (especially in the case of (114a)), the large contrast with the b-sentences is of importance.

(113) a. % Dat is het meisje **wat** ik denk **die** het gedaan heeft.
that is the girl what I think RP it done has
'That is the girl who I think has done it.' [72/255=28%]

b. ?* Dat is het meisje **die** ik denk **wat** het gedaan heeft.
that is the girl RP I think what it done has
[39/255=15%, MPQ2-A data]

(114) a. % Dat is het boek **wat** ik denk **die** de prijs gewonnen heeft.
that is the book what I think RP the prize won has
'That is the book that I think won the prize.' [47/255=18%]

b. ?* Dat is het boek **die** ik denk **wat** de prijs gewonnen heeft.
that is the book RP I think what the prize won has
[11/255=4%, MPQ2-A data]

There is however a caveat to this: a long-distance RC in which *wat* introduces the lower clause, whereas the higher clause is introduced by another element, is not unattested. More specifically, *wat* and *dat* seem to be interchangeable in long-distance RCs with a neuter gender antecedent – just like *die* and *wie* are in RCs with a common gender human antecedent (cf. section 2.3).[89] This is illustrated in (115) and (116).

(115) a. % Dat is het meisje **wat** ik denk **dat** het gedaan heeft.
that is the girl what I think that it done has
'That is the girl who I think has done it.' [128/255=50%]

b. % Dat is het meisje **dat** ik denk **wat** het gedaan heeft.
that is the girl that I think what it done has
'That is the girl who I think has done it.'
[66/255=26%, MPQ2-A data]

(116) a. % Dat is het boek **wat** ik denk **dat** de prijs gewonnen heeft.
that is the book what I think that the prize won has
'That is the book that I think won the prize.' [123/255=48%]

b. % Dat is het boek **dat** ik denk **wat** de prijs gewonnen heeft.
that is the book that I think what the prize won has
'That is the book that I think won the prize.'
[120/255=47%, MPQ2-A data]

These sentences can be accounted for in terms of spell out of multiple copies: *wat* and *dat* are equally suited to spell out the relative pronoun to a neuter gender (human/non-human) antecedent. This is abstractly illustrated in (117); recall that neuter is the underspecification for gender ([gen(der):]), and that both *wat* and *dat* are underspecified for gender, cf. section 2.5.3 for details.

(117) RC head$_{[GEN:\ ]}$ [$_{RC}$ DP$_{[OP,[GEN:\ ]]}$ … [$_{CP}$ DP$_{[OP,[GEN:\ ]]}$ …
dat*/*wat ***dat*/*wat***
long-distance restrictive relative clause

[89] It might be the case that the element *dat* in the sentences in (115) and (116) is in fact a complementizer instead of a pronoun. Since it is impossible to distinguish between the two options in the context of these sentences, I disregard the option of *dat* being a complementizer here, but see chapter 4 for discussion on the relation between pronoun *dat* and complementizer *dat*.

Even though the doubling patterns in (115) and (116) can easily be accounted for by a multiple copy spell out analysis, the occurrence of these sentences raises doubts about the correctness of the proposed analysis of the contrast in (113) and (114). That is to say, it is unclear whether the a-sentences in (113) and (114) really involve operator movement and double spell out, or whether they represent the realization of multiple copies that happen to have different surface forms (just as is the case for (115) and (116)). And if the latter is true, it is unclear why the b-sentences in (113) and (114) are ungrammatical. In long-distance RCs with a neuter gender antecedent that independently allow *wat* as a relative pronoun, it thus seems impossible to determine whether we are dealing with multiple copy spell out or with operator movement and double spell out.

However, the contrast between the patterns in (113) and (114) and the patterns in (115) and (116) might be explained by invoking the difference between spelling out syntactic gender and spelling out semantic animacy. Suppose that in the case of multiple copy spell out within a single movement chain, the copies can either spell out syntactic gender, or they can spell out semantic animacy, but it is impossible for one copy to spell out syntactic gender and for another copy to spell out semantic animacy. Now given the assumption that in the context of a neuter gender RC head, *wat* and *dat* spell out gender, whereas *die* spells out animacy (in accordance with the Individuation Hierarchy),[90] we can account for the contrast between (113) and (114) on the one hand, and (115) and (116) on the other hand. The doubling patterns in (113) and (114) cannot be the result of spelling out multiple copies because one copy would be the spell out of gender (*wat*), whereas the other copy would be the spell out of animacy (*die*). These sentences thus have to involve operator movement and double spell out, which explains the ungrammaticality of the b-sentences (they violate the *Inclusiveness Condition*). In contrast, the doubling patterns in (115) and (116) can be the spell out of multiple copies in a movement chain (both spell out gender), hence the grammaticality of both patterns (*wat-dat* and *dat-wat*).

2.6.2 Multiple embeddings

The proposed analysis of doubling predicts that in long A-bar dependencies with an extra embedding, the left periphery of all lower clauses can be introduced by a pronoun. This prediction is borne out by the MPQ2-B data, as illustrated in (118) for a long-distance root *wh*-Q with two embeddings.[91]

[90] Although this holds for a neuter gender *human* antecedent like *meisje* 'girl', it is somewhat of an oversimplification for a neuter gender *non-human* noun like *boek* 'book'. In the latter case, *dat* and *wat* may also be the spell out of semantic animacy (cf. table 2.11 in section 2.5.3 for details).

[91] The same goes for long-distance RCs with a common gender human antecedent with an additional embedding. All logically possible patterns with *die* and *wie* are attested in the MPQ2-B data (*w-w-w*, *w-w-d*, *w-d-d*, *w-d-w*, *d-d-d*, *d-d-w*, *d-w-w*, *d-w-d*), but the number of

(118) a. % **Wie** denk je **wie** Jan zei **wie** het gedaan heeft?
who think you who Jan said who it done has
'Who do you think Jan said has done it?' [149/380=39%]

b. % **Wie** denk je **wie** Jan zei **die** het gedaan heeft?
who think you who Jan said RP it done has
'Who do you think Jan said has done it?' [137/380=36%]

c. % **Wie** denk je **die** Jan zei **die** het gedaan heeft?
who think you RP Jan said RP it done has
'Who do you think Jan said has done it?' [103/380=27%]

d. % **Wie** denk je **die** Jan zei **wie** het gedaan heeft?
who think you RP Jan said who it done has
'Who do you think Jan said has done it?'
[106/380=28%, MPQ2-B data]

Notice that the observation that all variants of the construction in (118a) are attested – as illustrated in (118b)-(118d) – provides additional evidence for the claim that *die* and *wie* are interchangeable in the lower clause(s) of long-distance *wh*-Qs. Further empirical support for the claim that *die* and *wie* are interchangeable more generally, comes from coordinate structures. Assuming that coordination combines expressions of the same kind, we predict that in the case of two coordinate RCs one may be introduced by *die* and the other by *wie* (and vice versa).[92] This prediction is borne out by the MPQ2-B data, as illustrated in (119c) and (119d).

(119) a. de man **die** het gedaan heeft en **die** vervolgens is weggegaan
the man RP it done has and RP subsequently is gone
'the man who did it and subsequently left' [370/380=97%]

b. % de man **wie** het gedaan heeft en **wie** vervolgens is weggegaan
the man who it done has and who subsequently is gone
'the man who did it and subsequently left' [194/380=51%]

c. % de man **die** het gedaan heeft en **wie** vervolgens is weggegaan
the man RP it done has and who subsequently is gone
'the man who did it and subsequently left' [243/380=64%]

d. % de man **wie** het gedaan heeft en **die** vervolgens is weggegaan
the man who it done has and RP subsequently is gone
'the man who did it and subsequently left'
[256/380=67%, MPQ2-B data]

In a long-distance root *wh*-Q with two embeddings, subextraction of the operator (if any) is predicted to take place from either the most deeply embedded SpecCP or the intermediate SpecCP. This prediction is borne out, as illustrated by the sentences in (120) for *wh*-Qs involving the subextracted operator that is

attestations of these sentences is lower than the number of attestations of the long-distance *wh*-Qs with an additional embedding in (118) (all between 15% and 23%).

[92] Thanks to Adam Szczegielniak for pointing this out to me.

spelled out as *wat* and the A-bar pronoun *wie*. Sentence (120a) shows subextraction from the most deeply embedded SpecCP, and sentence (120b) shows subextraction from the intermediate SpecCP. As the left periphery of finite embedded clauses in Dutch needs to be introduced by at least one overt element (cf. section 2.2.1), the complementizer *dat* appears in absence of a pronoun.[93]

(120) a. % **Wat** denk je dat Jan zei **wie** het gedaan heeft?
what think you that Jan said who it done has
'Who do you think Jan said has done it?' [256/380=67%]

b. % **Wat** denk je **wie** Jan zei dat het gedaan heeft?
what think you who Jan said that it done has
'Who do you think Jan said has done it?'
[147/380=39%, MPQ2-B data]

Furthermore, it is predicted that in long-distance A-bar dependencies with multiple embeddings, the two strategies to form doubling constructions – namely subextraction plus double spell out and multiple copy spell out (of the pronoun or the operator) – can be combined. For the long-distance root *wh*-Qs in (120) this means that the left periphery of the clause in which *dat* surfaces should be able to be spelled out by a pronoun instead. This prediction is borne out as well, as illustrated in (121). Sentence (121a) illustrates subextraction from the most deeply embedded SpecCP (and double spell out as in (120a)) and additional spell out of the copy of the operator in the intermediate position. Sentence (121b) illustrates subextraction from the intermediate SpecCP (and double spell out as in (120b)) and additional spell out of the copy of the pronoun in the most deeply embedded SpecCP.[94]

(121) a. % **Wat** denk je **wat** Jan zei **wie** het gedaan heeft?
what think you what Jan said that it done has
'Who do you think Jan said has done it?' [183/380=48%]

b. % **Wat** denk je **wie** Jan zei **wie** het gedaan heeft?
what think you who Jan said who it done has
'Who do you think Jan said has done it?'
[139/380=37%, MPQ2-B data]

The proposed analysis thus correctly predicts that the highest clause or the highest clause *and* the intermediate clause can be introduced by *wat*. What

[93] As predicted by the analysis, the sentences in (120) with *die* instead of *wie* (i.e. *wat-dat-die* and *wat-die-dat*) are attested as well. More specifically, these sentences are accepted by 40% (153/380) and 20% (75/380) of the informants respectively (MPQ2-B data).

[94] Variants of the sentences in (121) in which (one of the instances of) *wie* is replaced by *die* are all attested (albeit it rather marginally in some cases): *wat-wat-die* (140/380=37%), *wat-die-die* (59/380=16%), *wat-wie-die* (148/380=39%), and *wat-die-wie* (69/380=18%). Notice that the occurrence of the latter doubling pattern is problematic for the analysis of BKL (cf. section 2.2.4): *wie* is assumed to be a subpart of *die*, so when both of them surface in a single movement chain, it should never be possible to find *die* in a higher clause than *wie*, as that would constitute a violation of the *Inclusiveness Condition*.

should be impossible is the appearance of *wat* in only the intermediate position (*wie-wat-wie*), because that would constitute a violation of the *Inclusiveness Condition*: after subextraction from the most deeply embedded SpecCP, features and structure are added during the last step of A-bar movement (from the intermediate SpecCP to the highest SpecCP). Put differently, as soon as we find an instance of *wat* in long-distance A-bar dependencies with two or more embeddings, we should find *wat* (or the declarative complementizer *dat* 'that' in intermediate SpecCPs) all the way up, i.e. it should be impossible to find a pronoun like *wie* introducing a higher clause than *wat*. A *wh*-Q with two embeddings that features the sequence *wat-wie-wat* was unfortunately not explicitly tested in the MPQ2, but it most likely does not occur, because the sequence *wat-wie* does not (or only very marginally, cf. section 2.2.1) occur in long-distance *wh*-Qs in the first place. Future research should settle this issue.

2.6.3 The ban on doubling with complex *wh*-phrases

In Standard Dutch, a long-distance *wh*-Q involving a complex *wh*-phrase that is introduced by *welke* 'which' looks as in (122). The complex *wh*-phrase itself introduces the higher SpecCP, whereas the lower CP domain is introduced by the invariant complementizer *dat* 'that'.[95]

(122) **Welke man** denk je **dat** het gedaan heeft?
which man think you that it done has
'Which man do you think has done it?'

It is well known that whereas doubling with pronouns (or doubling with PPs containing them, cf. section 2.7.2) occurs in many languages, doubling with complex *wh*-phrases is (practically) non-existent, as illustrated in (123) for Dutch.[96]

[95] I will illustrate the ban on doubling involving a complex *wh*-phrase introduced by *welke* 'which' only for *wh*-Qs, as it is independently impossible for such a complex *wh*-phrase to occur inside a restrictive RC (cf. Haeseryn et al. 1997:331, de Vries 2004:200, and see chapter 3 for discussion). However, let me point out here that the ban on doubling with complex *wh*-phrases is also attested in possessive RCs like the morphological genitive in (ia), as illustrated in (ib). The analysis to be proposed in this section covers this case as well. See also section 2.7.1 for non-identical doubling involving complex *wh*-phrases such as *wiens moeder* 'whose mother'.

(i) a. % de man **wiens moeder** jij denkt **dat** het gedaan heeft
the man whose mother you think that it done has
'the man whose mother you think has done it'

b. * de man **wiens moeder** jij denkt **wiens moeder** het gedaan heeft
the man whose mother you think whose mother it done has

[96] Alber (2008) shows that in Tyrolean, doubling with complex DPs is possible to some extent, as illustrated in (i). Based on these data she argues that doubling is sensitive to the heaviness of the extracted element rather than to the XP vs. X^0 status of the doubled element. This is more in line with the proposal outlined in the main text – according to which the application domain of *morphological merger* is not restricted to X^0 but may also

(123) * **Welke man** denk je **welke man** het gedaan heeft?
which man think you which man it done has
INTENDED: 'Which boy do you think has done it?'

From the proposal outlined in this chapter the ban on doubling with complex *wh*-phrases, as illustrated in (123), does not follow immediately.[97,98] That is to say, following Nunes (2004) in taking doubling to be the result of morphological reanalysis between the pronoun and the C head – as a result of which multiple copies can be spelled out – and following BKL in assuming that morphological reanalysis may target a phrase (XP) as well (instead of only an X^0), it is unclear why complex *wh*-phrases cannot be spelled out multiple times.[99]

To account for the ungrammaticality of doubling with complex *wh*-phrases, I adopt a proposal by van Craenenbroeck (2004, 2010) who argues that complex *wh*-phrases are *base-generated* in the left periphery whereas *wh*-pronouns *move* to the left periphery. It is a well known fact that bare pronouns behave syntactically differently from complex *wh*-phrases. Whereas Pesetsky (1987) attributes this difference to *D(iscourse)-linking* – bare pronouns being non D-linked and

apply to XPs – than with Nunes' (2004) proposal, cf. section 2.2.3. Whether or not the complex *wh*-phrases in Tyrolean that allow doubling have properties that are different from the properties of complex *wh*-phrases in Dutch that do not allow doubling is an issue I leave for future research.

(i) a. ?? **Prum** glapsch du, **prum** dass dr Hons net kemmen isch?
why think you why that the Hons not come is
'Why do you think Hons did not come?'

b. ?? die Fraindin, **mit der** wos sie glap, **mit der** wos die M. spieln tat
the friend with RP C she thinks with RP C the M. play would
'the friend with which she thinks that Maria would play'

c. ?? 's Madl, **wegn den** wos sie glap, **wegn den** wos die M. net kimp
the girl because RP C she thinks because RP C the M. not comes
'the girl because of which she thinks that Maria does not come'

d. ?* **Wellawegn** glapsch du, **wellawegn** dass dr Hons net kemmen isch?
why think you why that the Hons not come is
'Why do you think Hons did not come?'

e. ?* der Pua, **in Votr von den** wos i glaap, **in Votr von den** wos i gsechn hon
the boy the father of RP C I think the father of RP C I seen have
'the boy the father of which I think I have seen' [Alber 2008:150]

[97] If *all* doubling were the result of subextraction (but see footnote 109), the ban on doubling with complex *wh*-phrases would follow automatically. That is to say, since subextraction targets the operator in SpecDP *to the exclusion of the lexical NP*, there is no way for the lexical NP (*man* in (123)) to end up in the higher CP in a doubling configuration.

[98] An interesting line of thought about the ban on doubling with *wh*+NP phrases is pursued by Rett (2006). She argues that *wh*-phrases without an NP complement are non-quantificational, whereas *wh*-phrases with an NP complement are quantificational. The latter type cannot be copied as interpreting these phrases twice in the derivation leads to vacuous quantification. Notice that this line of thought requires there to be a direct relation between pronunciation and interpretation, i.e. pronouncing a copy twice has direct consequences for the interpretation of a construction.

[99] The possibility of PP doubling *does* immediately follow from this proposal, as will be shown in section 2.7.2.

*which NP*s being D-linked – van Craenenbroeck (2004, 2010) attributes this difference to the (non-)operator status of the elements involved. More specifically, he argues that minimal *wh*-phrases (bare *wh*-pronouns and PPs containing them) behave differently from complex *wh*-phrases in the sense that the former but not the latter act as operators.[100] Consequently, *wh*-pronouns *move* to the left periphery, whereas *wh*-phrases are *base-generated* in the left periphery while a coindexed *empty operator* moves to the left periphery (notice that this requires a split CP domain, cf. section 3.4.1 for more details). For a complete overview of the arguments in favor of this claim, I refer the reader to van Craenenbroeck (2004, 2010). Here I only mention a few of them. First, quantifiers, as opposed to referential expressions, cannot occur in Contrastive Left Dislocation (CLD) constructions in Dutch, as illustrated in (124).

(124) a. [Die jongens]$_i$, die$_i$ ken ik niet t_{die}.
those boys DEM know I not
'Those boys, I don't know.'

b. * Iedereen$_i$, die$_i$ ken ik niet t_{die}.
everybody DEM know I not [van Craenenbroeck 2004:37]

Similarly, complex *wh*-phrases may marginally occur in CLD constructions (125a) – in contrast to bare *wh*-pronouns (125b) – which seems to suggest that complex *wh*-phrases do not function as syntactic operators (quantifiers).[101,102]

(125) a. ?? [Welke jongen]$_i$, die$_i$ heb je t_{die} gezien?
which boy DEM have you seen

b. * Wie$_i$, die$_i$ heb je t_{die} gezien?
who DEM have you seen [van Craenenbroeck 2004:37]

A second argument in favor of the claim that complex *wh*-phrases do not function as operators comes from preposition stranding in Dutch. Van Riemsdijk (1978) observed that prepositions in Dutch may only be stranded by R-pronouns or empty operators. As illustrated in (126), an R-pronoun (here *waar*)

[100] Semantically, it seems to make sense to assume that complex *wh*-phrases like *welke man* 'which man' or *wiens moeder* 'whose mother' do not act as operators in their entirety. If a sentence like *Wiens moeder heeft het gedaan?* 'Whose mother has done it?' gets answered with *Mary*, this *Mary* does not refer to the *wh*-phrase *wiens moeder* as a whole, rather only to the *wiens* part: *Mary's mother*. Syntactically, it seems to make sense as well: the operator is too deeply embedded inside the complex *wh*-phrase to make the whole *wh*-phrase act as an operator.

[101] As noted by van Craenenbroeck (2004:38), the fact that (125a) is not completely grammatical is most likely due to the fact that CLD requires purely referential DPs, which *welke jongen* 'which boy' is not.

[102] Both sentences in (125) were explicitly tested in MPQ2-A, but they were not attested: only 1%-3% of the informants indicated that these sentences occur in their spoken Dutch. A reason for the lack of attestations of sentence (125a) in MPQ2-A might be related to the fact that this sentence requires a very specific intonation pattern for it to be (somewhat) acceptable (no comma intonation). This intonation pattern was not completely well presented to the informants of MPQ2-A.

can strand a preposition – in a *wh*-Q (126a) and in a RC (126b) – whereas its non R-pronoun counterparts cannot (here *wat* or *die*). The sentences in (127) show that in constructions that arguably involve operator movement – topic drop as in (127a) and an infinitival purpose clause as in (127b) – the preposition can be stranded.

(126) a. [*Waar*/ **wat*] heb je die kist ***mee*** opengemaakt?
where what have you that crate with open made
'What did you open that crate with?'

b. de koevoet [*waar*/ **die*] ik de kist ***mee*** opengemaakt heb
the crowbar where RP I the crate with open made have
'the crowbar I opened the crate with'
[van Craenenbroeck 2004:40]

(127) a. *Op* heb ik al ***mee*** gewerkt.
have I already with worked
'I have already worked with that.'

b. Die sleutel is te klein [*Op* om het slot ***mee*** open te doen].
that key is too small for the lock with open to do
'That key is too small to open the lock with.'
[van Craenenbroeck 2004:40]

There is a contrast between bare *wh*-pronouns and complex *wh*-phrases with respect to preposition stranding: whereas the first cannot strand a preposition, the latter may successfully do so, as illustrated in (128). If we want to maintain van Riemsdijk's (1978) generalization, this suggests that only questions with complex *wh*-phrases involve empty operator movement.[103]

(128) a. * Wie wil je niet mee samenwerken?
who want you not with cooperate
INTENDED: 'Who don't you want to cooperate with?'

b. ? Welke jongen wil je niet mee samenwerken?
which boy want you not with cooperate
'Which boy don't you want to cooperate with?'
[van Craenenbroeck 2010:249]

[103]This is somewhat of an oversimplification of matters. The sentences in (128) were tested in MPQ2-A, and although the b-sentence is indeed attested more frequently than the a-sentence, both sentences are attested rather frequently in the Dutch speaking language area: respectively 59% (151/255) and 78% (198/255) of the informants accept these sentences. It is well known that bare *wh*-pronouns seem to behave syntactically like complex *wh*-phrases in certain D-linked contexts (cf. Pesetsky 1987), i.e. in the right context, a *wh*-pronoun (which is normally not D-linked) may behave as if it were D-linked. As there was no control for context in the MPQs, it might have been the case that the informants that accepted the sentence in (128a) interpreted the *wh*-pronoun as being D-linked, hence on a par with a complex *wh*-phrase. In terms of van Craenenbroeck's approach to *wh*-pronouns and complex *wh*-phrases – according to which the effect of D-linking is epiphenomenal to structural complexity – this means that for some informants in certain contexts (here (128a)) the 'implicit N-restriction' of a *wh*-pronoun can become syntactically accessible (van Craenenbroeck 2004:47), as a result of which the bare *wh*-pronoun will behave syntactically similar to a complex *wh*-phrase.

From van Craenenbroeck's (2004, 2010) analysis of the difference between complex *wh*-phrases and bare *wh*-pronouns, the lack of doubling with complex *wh*-phrases immediately follows: there is no movement chain of the complex *wh*-phrase, so there is not more than one copy that can be spelled out.

2.6.4 Intervention effects

Recall from section 2.2.2 that identical and non-identical doubling in long-distance *wh*-Qs pattern alike with respect to intervening negation. The relevant examples are repeated here in (129) (=(24)).

(129) a. **Wie** denk je *niet* dat zij uitgenodigd heeft?
who think you not that she invited has
'Who don't you think that she invited?'

b. * **Wat** denk je *niet* **wie** zij uitgenodigd heeft?
what think you not who she invited has

c. * **Wie** denk je *niet* **wie** zij uitgenodigd heeft?
who think you not who she invited has [BKL 2009:40]

More or less the same pattern is attested for intervening negation in long-distance RCs, as illustrated in (130); even though (130b) is accepted by more than 15% of the informants (hence the % in front of it), it is accepted considerably less frequently than (130a) and thus seems to be more on a par with (130c). The reason why not all informants accept sentence (130a) – in contrast to the *wh*-Q in (129a) – is most likely that there is an alternative for forming long-distance RCs: *resumptive prolepsis*, as illustrated in (131) (cf. section 2.3). This sentence is accepted by 95% of the informants.[104]

(130) a. % Dat is het meisje **dat/wat** ik *niet* denk dat het gedaan heeft.
that is the girl that/what I not think that it done has
'That is the girl who I don't think has done it.'
[101/255=40%, 87/255=34%]

b. % Dat is het meisje **wat** ik *niet* denk **wat** het gegaan heeft.
that is the girl what I not think what it done has
[44/255=17%]

c. ?* Dat is het meisje **wat** ik *niet* denk **die** het gedaan heeft.
that is the girl what I not think RP it done has
[33/255=13%, MPQ2-A data]

(131) Dat is het meisje **waarvan** ik *niet* denk **dat** zij het gedaan heeft.
that is the girl whereof I not think that she it done has
'That is the girl who I don't think has done it.'
[243/255=95%, MPQ2-A data]

[104]The absence of intervention effects in the resumptive prolepsis construction suggests that it indeed differs from the long-distance A-bar movement constructions discussed in this chapter (cf. section 2.3).

The observation that identical and non-identical doubling (roughly) pattern alike with respect to intervening negation suggests that they should receive a similar analysis. These facts are thus compatible with my analysis according to which both identical and non-identical doubling are the result of successive-cyclic movement of (part of) the A-bar pronoun and multiple copy or double spell out. More specifically, by relying on Pesetsky's (2000) notion of *intervention effect* as given in (132), the patterns in (129) and (130) are predicted under my analysis of doubling. Assuming that in doubling configurations, the higher copy is interpreted as the operator and the lower copy is interpreted as the restriction (cf. BKL 2009:40), the contrast between the a-sentences in (129) and (130) on the one hand and the b- and c-sentences in (129) and (130) on the other hand is explained: only in the b- and c-sentences there is an intervening scope-bearing element, as a result of which only those sentences are ungrammatical (or at least degraded).

(132) *Intervention effect* (Pesetsky 2000:67)
A semantic restriction on a quantifier (including *wh*) may not be separated from that quantifier by a scope-bearing element

However, as noted by BKL and as is also shown by the MPQ2-A data, things become less clear when an intervening universal quantifier is taken into account. As illustrated in (133), doubling configurations in *wh*-Qs are insensitive to an intervening quantifier, i.e. unlike with negation, there is no difference between non-doubling (133a) and doubling constructions (133b,c). The only difference between (133a) and (133b,c) is the observation that the latter are not accepted by all speakers (colloquial Dutch vs. Standard Dutch), just like their counterparts without an intervening quantifier (cf. section 2.2).

(133) a. **Wie** denkt *iedereen* dat een goede president is geweest?
who thinks everyone that a good president is been
'Who does everyone think was a good president?'

b. % **Wie** denkt *iedereen* **wie** een goede president is geweest?
who thinks everyone who a good president is been

c. % **Wat** denkt *iedereen* **wie** een goede president is geweest?
what thinks everyone who a good president is been

[BKL 2009:40-41]

As illustrated in (134), for RCs the pattern is roughly the same; only the status of identical doubling with respect to an intervening universal quantifier (134b) is a bit unclear: it is accepted by merely 18% of the informants. Notice that all informants allow the resumptive prolepsis construction with an intervening universal quantifier as in (135), which might explain the fact that not all informants allow the construction in (134a), unlike the *wh*-Q in (133a).

(134) a. % Dat is het meisje **dat/wat** *iedereen* denkt dat het gedaan heeft.
that is the girl that/what everyone thinks that it done has
'That is the girl everyone thinks has done it.
[100/255=39%, 85/255=33%]

b. % Dat is het meisje **wat** *iedereen* denkt **wat** het gedaan heeft.
that is the girl what everyone thinks what it done has
[45/255=18%]

c. % Dat is het meisje **wat** *iedereen* denkt **die** het gedaan heeft.
that is the girl what everyone thinks RP it done has
[71/255=28%, MPQ2-A data]

(135) Dat is het meisje **waarvan** *iedereen* denkt **dat** zij het gedaan heeft.
that is the girl whereof everyone thinks that she it done has
'That is the girl everyone thinks has done it.'
[252/255=99%, MPQ2-A data]

One possible way of accounting for the differences between intervention effects with negation and intervention effects with a universal quantifier might be by means of Quantifier Raising (QR, LF-raising of quantifiers), i.e. QR of the universal quantifier obviates the intervention effect in (133b,c) and (134b,c).[105] As negation is not subject to QR, the intervention effect in (129b,c) and (130b,c) cannot be obviated in a similar vein (but see BKL 2009:41-43 for a discussion of the problems that such a theory faces).

These results regarding intervention effects are obviously preliminary and subject to further empirical investigation. As witnessed by the unclear status of the sentences in (130b) and (134b), differences in the status of intervention effects can be very subtle (cf. also appendix A to this chapter). For this reason, I leave the status and nature of intervention effects to future research.

2.7 Extensions

This section expands the data set to include other instances of doubling in long-distance A-bar dependencies. Starting from the analysis of pronominal doubling as presented in the preceding sections, I sketch the outlines of a unified theory of doubling that covers doubling involving complex *wh*-phrases (section 2.7.1) and doubling with prepositional phrases containing an A-bar pronoun (section 2.7.2) as well. In section 2.7.3, I briefly mention subject clitic doubling within the clause, and the analysis of this phenomenon as proposed by van Craenenbroeck and van Koppen (2008). The similarity between their analysis and mine suggests that a unified account of doubling – i.e. doubling across clause boundaries and doubling within the clause – is within reach.

[105] I leave open here the question of the landing site targeted by quantifier raising.

2.7.1 Subextraction from complex *wh*-phrases

As already mentioned in section 2.6.3, Standard Dutch long-distance A-bar dependencies involving a complex *wh*-phrase look as in (136) and (137): the complex *wh*-phrase introduces the higher clause, whereas the lower clause is introduced by the invariant finite declarative complementizer *dat* 'that'.[106]

(136) **Welke man** denk je **dat** ik gisteren gezien heb?
which man think you that I yesterday seen have
'Which man do you think I have seen yesterday?'

(137) a. % de man **wiens moeder** jij denkt **dat** het gedaan heeft
the man whose mother you think that it done has
'the man whose mother you think has done it' [153/255=60%]

b. % het meisje **wiens moeder** jij denkt **dat** het gedaan heeft
the girl whose mother you think that it done has
'the girl whose mother you think has done it'
[150/255=59%, MPQ2-A data]

Variants of the constructions in (136) and (137) that occur in colloquial Dutch are given in (138) (=(44)) for *wh*-Qs and in (139) and (140) for RCs: the complex *wh*-phrase remains in the lower CP domain, whereas an A-bar pronoun surfaces in the higher CP domain.[107] For *wh*-Qs this A-bar pronoun needs to be a *wh*-pronoun: *wat* or *wie*. For RCs, this pronoun can be any A-bar pronoun that is compatible with the given RC head: *wie* or *die* with the common gender human RC head *man* 'man', and *wat*, *wie* or *die* with the neuter gender human RC head *meisje* 'girl'.

[106]The observation that not all informants accept the possessive RCs in (137) is most likely related to the fact that Dutch has more than one way to form a possessive RC, as illustrated in (i) (cf. also section 4.2.2).

(i) a. de man wiens vader ik ken
the man whose father I know
'the man whose father I know' [morphological genitive]

b. de man wie z'n vader ik ken
the man who his father I know
'the man whose father I know' [relative plus possessive pronoun]

c. de man van wie ik de vader ken
the man of who I the father know
'the man whose father I know' [prepositional genitive, de Vries 2006:2]

In addition, even though *wiens* is the genitive form that is used with all antecedents in informal speech, *wier* is used with feminine and plural antecedents in formal speech (cf. ANS, Haeseryn et al. 1997:343). So, part of the reason for why not all speakers accept (137b) might be that some speakers prefer *wier* over *wiens* in this context.

[107]The reverse doubling patterns of the patterns in (138), (139) and (140), namely doubling patterns in which the higher clause is introduced by a complex *wh*-phrase whereas the lower clause is introduced by an A-bar pronoun, are also attested in colloquial Dutch (albeit less frequently). I will discuss such patterns in appendix B to this chapter, as these data strictly speaking fall outside the scope of the analysis proposed in this chapter.

(138) a. % **Wat** denk je **welke man** ik gisteren gezien heb?
what think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'
[240/333=72%]

b. % **Wie** denk je **welke man** ik gisteren gezien heb?
who think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'
[211/333=63%, MPQ1-B data]

(139) a. ?* de man **wat** jij denkt **wiens moeder** het gedaan heeft
the man what you think whose mother it done has
'the man whose mother you think has done it' [13/255=5%]

b. % de man **wie** jij denkt **wiens moeder** het gedaan heeft
the man who you think whose mother it done has
'the man whose mother you think has done it' [79/255=31%]

c. % de man **die** jij denkt **wiens moeder** het gedaan heeft
the man RP you think whose mother it done has
'the man whose mother you think has done it'
[103/255=40%, MPQ2-A data]

(140) a. % het meisje **wat** jij denkt **wiens moeder** het gedaan heeft
the girl what you think whose mother it done has
'the girl whose mother you think has done it' [71/255=28%]

b. % het meisje **wie** jij denkt **wiens moeder** het gedaan heeft
the girl who you think whose mother it done has
'the girl whose mother you think has done it' [77/255=30%]

c. % het meisje **die** jij denkt **wiens moeder** het gedaan heeft
the girl RP you think whose mother it done has
'the girl whose mother you think has done it'
[67/255=26%, MPQ2-A data]

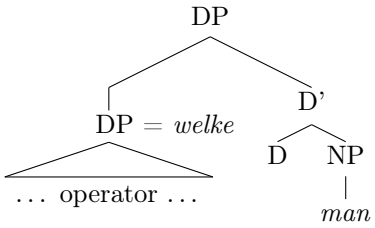
As for complex *wh*-phrases introduced by *welke* 'which', I follow a proposal by Leu (2008) and assume that *welk(e)* itself is the spell out of a phrase (contra e.g. Corver 1990, Longobardi 1994). Dutch *welk(e)* (and its cognates in other languages, cf. Leu 2008) is morphologically composed of a *w*-morpheme – comparable to the *w*-morpheme in the A-bar pronouns *wie* and *wat* – and an element *elk(e)* (which might be morphologically complex in itself). Besides being morphologically complex, evidence in favor of the claim that *welk(e)* spells out a phrase comes from the observation that *welke* may function as a relative pronoun in *restrictive* RCs in formal Dutch – just like the Dutch A-bar pronouns *die*, *dat*, *wie* and *wat* – suggesting that *welke* contains an operator in SpecDP as well. This is illustrated in (141).

(141) a. De procedures **welke** bij zo'n gelegenheid gevolgd worden,
the procedures which at such an occasion followed be
zijn verouderd.
are outdated
'The procedures that are followed on such an occasion are outdated.'

b. De klok **welke** ik u wil aanbieden, is honderd jaar oud.
the clock which I you want to offer is hundred years old
'The clock that I want to offer you is a hundred years old.'
[formal Dutch, Haeseryn et al. 1997:336]

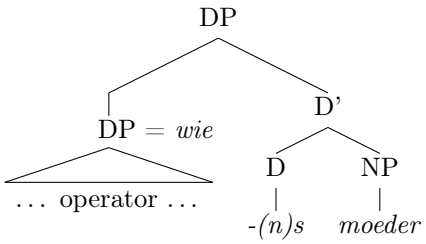
The syntactic structure of *welk(e) NP*, here illustrated for *welke man* 'which man', should thus look like (142).

(142)



It is irrelevant to the present discussion what analysis of possessive phrases is adopted. The important thing to note is that a phrase like *wiens moeder* 'whose mother' contains the A-bar pronoun *wie* 'who', that in turn contains an operator in SpecDP, just as is the case with a complex *wh*-phrase introduced by *welke* 'which' (142). This is indicated for the phrase *wiens moeder* by the simplified structure in (143).

(143)



Recall that I followed van Craenenbroeck (2004, 2010) in assuming that complex *wh*-phrases as a whole do not act as operators, but are base-generated in the left periphery (as a result of which complex *wh*-phrases cannot be doubled, cf. section 2.6.3). Movement of the whole complex *wh*-phrase is thus impossible, and we are left with the following two possibilities: either the operator subextracts from the A-bar pronoun, or the operator pied pipes the A-bar pronoun along with it. However, pied piping of the A-bar pronoun containing the operator leads to a violation of the Left Branch Condition (LBC, Ross 1967:207),

according to which extraction of a noun phrase on the left branch of another noun phrase leads to ungrammaticality. This is illustrated in (144). In light of the earlier proposed operation of *rescue by PF spell out*, the question immediately arises as to why the LBC violation in (144) cannot be ameliorated by PF spell out (i.e. **welke-welke man*). As mentioned in section 2.5.5, PF spell out can only apply in case deletion of the offending copy leads to a recoverability problem (deletion upon recoverability); recall that PF deletion is more economical than PF spell out. Whereas a recoverability problem arises in case the operator is subextracted from an A-bar pronoun and the latter is deleted, there is no recoverability problem is case a full A-bar pronoun is subextracted from a complex *wh*-phrase and the lower instance of the copy of the A-bar pronoun is deleted. So, only in the former but not in the latter case may PF spell out obviate a locality violation.

(144) * **Welke** denk je **man** ik gisteren gezien heb?
which think you man I yesterday seen have
INTENDED: 'Which man do you think I have seen yesterday?'

The only option is thus to subextract the operator from the A-bar pronoun and move it to the higher left periphery, as abstractly illustrated in (145).

(145) $[_{CP}$ operator$_1$ ... $[_{CP}$ [pronoun$_1$ NP] ...
subextraction

The operator is spelled out by *wat* – *wat* being the most underspecified A-bar pronoun in Dutch (cf. *supra*) – giving rise to the structures in (138a) and (140a). This is exemplified in (146).

(146) $[_{CP}$ operator$_1$ ... $[_{CP}$ $[_{DP}$ pronoun$_1$ NP] ...
wat ***welke man/wiens moeder***

As for the *wh*-Q in (138b), in which the higher pronoun surfaces as *wie*, I would like to speculate – building on a proposal by den Dikken (2009b) – that the operator in SpecDP may enter into a *concord* relation (i.e. agreement/feature sharing) with some (or all) of the features of DP before it moves out of the DP (subextraction).[108] The result of this concord is that the subextracted operator may receive a different spell out than *wat*. In the context of a *wh*-Q questioning a person, the operator may enter into a concord relation with the [human] feature, as a result of which it will be spelled out as *wie*. More specifically, the Closest Match Principle selects the Vocabulary Item *wie* as the optimal realization of this concordial operator, because it matches most features of the operator to be lexicalized (*wie* furthermore satisfies the *wh*-requirement on the

[108]This type of concord is different from the concord proposed by den Dikken (2009b) in the sense that the elements that share features originated within the same DP. Put differently, whereas in den Dikken's proposal two separate items share features, in my proposal feature sharing happens between two elements that originated within the same constituent.

introduction of *wh*-Qs). That is to say, whereas Vocabulary Item *wat* may in principle lexicalize the concordial operator, it loses the competition for insertion from the more specific *wie*. This is illustrated in (147).

(147) $[_{CP[+wh]}$ operator$_{[HUMAN]}$ ... $[_{CP}$ $[_{DP}$ pronoun NP$]_{[HUMAN]}$ ...
wie ***welke man***

The same holds for the RC constructions in (139b,c) and (140b,c), in which the higher pronoun surfaces as *wie* or *die*. Before the operator moves out of the DP it may enter into a concord relation with the [human] feature, as a result of which it can be spelled out as *wie* or *die* (in accordance with the Individuation Hierarchy, cf. *supra* (82)), as illustrated in (148). Notice that this is only possible as long as the feature(s) on the operator do(es) not conflict with the features on the RC head, i.e. the features on the pronoun introducing the RC need to match the features on the RC head (cf. *supra*).

(148) $[_{CP}$ operator$_{[HUMAN]}$ ... $[_{CP}$ $[_{DP}$ pronoun NP$]_{[HUMAN]}$ ...
wie / ***die*** ***wiens moeder***

Needless to say, further research is necessary in order to elucidate more fully the properties of concord and the role it plays in doubling configurations.[109]

2.7.2 Doubling of prepositional phrases

Besides doubling of A-bar pronouns in long-distance A-bar dependencies, prepositional phrases containing an A-bar pronoun can also be doubled in colloquial Dutch, as illustrated in (149b) and (150b) for a *wh*-Q and a RC respectively. The a-sentences in (149) and (150) represent Standard Dutch: the highest clause is introduced by the prepositional phrase and the lower clause is introduced by the finite declarative complementizer *dat*.[110] Notice that doubling of prepositional phrases seems to indicate that identical doubling is indeed the result of the spell out of multiple copies, cf. section 2.2.3.

[109] Notice that by adding the notion of concord to the system of doubling, we get redundancy in the sense that a doubling configuration like *wie-wie* could be the result of multiple copy spell out as well as the result of subextraction, double spell out and concord. Put differently, it becomes hard (if not impossible) to distinguish two derivations for a single doubling pattern. In earlier work I therefore proposed that all doubling is the result of subextraction and double spell out, the difference between 'identical' and non-identical doubling being whether or not there is concord. The reason why I no longer pursue this approach is that it cannot account for doubling in long A-bar dependencies with an additional embedding (cf. section 2.6.2) without assuming a mechanism of multiple copy spell out (e.g. Nunes 2004). If such a mechanism is needed anyway, it seems no longer desirable to reduce all doubling to subextaction and double spell out as 'identical' doubling is most easily accounted for by multiple copy spell out. Furthermore, a unified account of doubling in terms of subextraction, double spell out and (optional) concord cannot easily account for doubling of prepositional phrases (cf. section 2.7.2) whereas a multiple copy spell out account can (see also appendix A to this chapter).

[110] The fact that not all informants accept the sentence in (150a) – whereas all informants accept the sentence in (149a) – is most likely due to the nature of this particular test sentence: it attributes a thought to someone. This might have influenced the rate of acceptance of this sentence, as well as of the sentences in (150b) and (152).

(149) a. **Op wie** denk je **dat** hij verliefd is?
on who think you that he in love is
'Who do you think he is in love with?' [378/380=99%]

b. % **Op wie** denk je **op wie** hij verliefd is?
on who think you on who he in love is
'Who do you think he is in love with?'
[189/380=50%, MPQ2-B data]

(150) a. % Dat is het meisje **op wie** je denkt **dat** hij verliefd is.
that is the girl on who you think that he in love is
'That is the girl I think he is in love with.' [300/380=79%]

b. % Dat is het meisje **op wie** je denkt **op wie** hij verliefd is.
that is the girl on who you think on who he in love is
'That is the girl I think he is in love with.'
[116/380=31%, MPQ2-B data]

Doubling of prepositional phrases containing an A-bar pronoun is correctly predicted to exist by the proposal outlined in this chapter. Recall that I assume that, just like A-bar pronouns themselves, prepositional phrases containing an A-bar pronoun act as operators (unlike complex *wh*-phrases, cf. section 2.6.3), and therefore *move* successive-cyclically from their base position to the higher left periphery. Now, as I assume that morphological reanalysis may target the C head and a *phrasal* element in SpecCP (cf. BKL) – i.e. not only *heads* are subject to the process of morphological reanalysis (*pace* Nunes 2004, cf. section 2.2.3) – nothing prohibits morphological reanalysis to target the C head and a prepositional phrase in SpecCP, as a result of which multiple copies of the prepositional phrase may be spelled out.

Interestingly, just like doubling with complex *wh*-phrases (cf. section 2.7.1), doubling involving prepositional phrases containing an A-bar pronoun seems to exhibit subextraction (and concord) as well, suggesting that my analysis is on the right track. More specifically, the operator inside the A-bar pronoun that is the complement of the preposition may subextract, thereby giving rise to constructions like (151a) and (152a) in case there is no concord, and to constructions like (151b) and (152b)-(152c) in case there is concord (i.e. the operator and the prepositional phrase share the feature [human]).[111,112,113]

[111]Notice that the doubling patterns in (152b) and (152c) could in principle also be analyzed as the result of subextraction of the A-bar pronoun, instead of subextraction of only the operator.

[112]For *wh*-Qs, the MPQ2-B data show that the same doubling patterns are attested when an R-pronoun is involved: *waarop-dat* 'whereon-that' (Standard Dutch, deletion of all but highest copy), *waarop-waarop* 'whereon-whereon' (colloquial Dutch, spell out of multiple copies), *wat-waarop* 'wat-whereon' (colloquial Dutch, subextraction and double spell out), and *waar-waarop* 'where-whereon' (colloquial Dutch, subextraction and double spell out plus concord). The same is not true for RCs: doubling constructions involving an R-pronoun are attested only marginally.

[113]The reverse patterns of the doubling constructions in (151) and (152) are attested as well, although such constructions are attested considerably less frequently than their prepositional

(151) a. % **Wat** denk je **op wie** hij verliefd is?
what think you on who he in love is
'Who do you think he is in love with?' [164/380=43%]

b. % **Wie** denk je **op wie** hij verliefd is?
who think you on who he in love is
'Who do you think he is in love with?'
[260/380=68%, MPQ2-B data]

(152) a. % Dat is het meisje **wat** je denkt **op wie** hij verliefd is.
that is the girl what you think on who he in love is
'That is the girl I think he is in love with.' [82/380=22%]

b. % Dat is het meisje **wie** je denkt **op wie** hij verliefd is.
that is the girl who you think on who he in love is
'That is the girl I think he is in love with.' [88/380=23%]

c. % Dat is het meisje **die** je denkt **op wie** hij verliefd is.
that is the girl RP you think on who he in love is
'That is the girl I think he is in love with.'
[99/380=26%, MPQ2-B data]

2.7.3 Subject clitic doubling

Several southern Dutch varieties display so-called *clitic doubling*, a phenomenon whereby a strong subject pronoun is doubled by a clitic pronoun (e.g. Haegeman 1992, 2004b, van Craenenbroeck and van Koppen 2002a,b, 2007, 2008). The two pronouns together form the subject of the sentence. Clitic doubling occurs in embedded clauses (153a) and inverted main clauses (153b).[114]

(153) a. da **ze** **zaa** gisteren gewerkt ee
that she$_{CLITIC}$ she$_{STRONG}$ yesterday worked has
'that she has worked yesterday'

doubling counterparts in (151) and (152); this is especially true for *wh*-Qs. I briefly discuss such doubling patterns in appendix B to this chapter.

[114] *Clitic doubling* needs to be distinguished from what van Craenenbroeck and van Koppen (2002a) call *topic doubling*: pronominal doubling in subject initial main clauses in which the first subject element cannot be a clitic, but must be a weak pronoun, a strong pronoun, a full DP, or a proper name. Van Craenenbroeck and van Koppen (2002a) argue that the first element in this construction is a topic that is base-generated in SpecCP, whereas the strong subject pronoun is the 'real' argument of the verb. As the two elements in *topic doubling* constructions arguably are not members of the same movement chain, I will disregard this type of pronominal doubling in the remainder of this section.

(i) <***Me** / **We**> gojn ze **waaile** nuir ojsh bringen.
we$_{CLITIC}$ we$_{WEAK}$ go them we$_{STRONG}$ to home bring
'We're going to take them home.'
[Wambeek Dutch, van Craenenbroeck and van Koppen 2002a:281]

b. Gisteren ee **ze** **zaa** gewerkt.
yesterday has she$_{CLITIC}$ she$_{STRONG}$ worked
'Yesterday she has worked.'
[Gent Dutch, van Craenenbroeck and van Koppen 2002a:281]

On the basis of the patterns found in clitic doubling constructions with coordinated subjects (cf. section 2.5.5), van Craenenbroeck and van Koppen (2008) argue in detail that in the relevant dialects, the clitic pronoun (pro-PhiP) is properly contained in the strong pronoun (pro-DP), as illustrated in (154).

(154) [$_{DP}$ = strong pronoun *zaa* [$_{PhiP}$ = clitic pronoun *ze* [NP ø]]]

Doubling comes about through subextraction of PhiP and double spell out: the extracted PhiP part of the structure (in SpecFinP) gets spelled out by a clitic and the DP part of the structure (in SpecTP) gets spelled out by a strong subject pronoun. From this analysis, the lack of clitic doubling with lexical NP subjects follows as well: because PhiP dominates/contains the NP that contains lexical material, this PhiP can never be spelled out as a clitic.

The analysis of subject clitic doubling as proposed by van Craenenbroeck and van Koppen (2008) is thus very similar to the analysis I propose for pronominal doubling in long A-bar dependencies: both assume subextraction (although the position/element that subextraction targets differs) and double spell out. This correspondence between the two accounts promises the possibility of extending the domain of application of the subextraction plus double spell out account of doubling from intra-clausal dependencies to dependencies within the clausal domain.

2.8 Conclusion

Building and improving on a proposal by Barbiers, Koeneman, and Lekakou (2009) on doubling in long-distance root *wh*-Qs in Dutch, in this chapter, I proposed a unified analysis of pronominal doubling in Dutch long-distance A-bar dependencies on the basis of new empirical data on doubling patterns in long-distance RCs. I assumed that all long A-bar dependencies are derived by successive-cyclic movement via SpecCP of (part of) the A-bar pronoun. At PF, either only the highest copy of the pronoun is spelled out (155a), or the highest copy *and* the intermediate copy of the pronoun in the embedded SpecCP are spelled out (155b). In syntax, either the full pronoun moves up, or only the operator – located in the specifier of the pronoun – moves up. The latter scenario results in spell out of the operator as well as spell out of the pronoun from which it extracted, for recoverability reasons and to salvage an otherwise illicit step in the derivation: *rescue by PF spell out* (155c).

(155) a. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ ~~pronoun$_1$~~ ... ~~pronoun$_1$~~ ...]]
no doubling

b. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
'identical' doubling

c. $[_{CP}$ **operator**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
non-identical doubling

The (differences in the) specific patterns of pronominal doubling (in *wh*-Qs and RCs) are best accounted for by the feature specifications of the A-bar pronouns involved (*d*-pronouns and *w*-pronouns), and different lexicalization options and requirements. That is to say, the exact distribution of *d*-pronouns and *w*-pronouns is related to the features they spell out: syntactic gender (common/neuter) and/or semantic animacy (roughly human/non-human). A late insertion model of morphology accounts for the fact that *d*-pronouns and *w*-pronouns are interchangeable in certain contexts (for some speakers), as both can be equally compatible with the structure to be lexicalized. Whether a *d*-pronoun or a *w*-pronoun will be inserted was furthermore shown to be dependent on the nature of the clause: *wh*-Q or RC. More specifically, *d*-pronouns cannot introduce *wh*-Qs due to the *wh*-requirement on the introduction of such clauses.

Doubling in RCs and doubling in *wh*-Qs in the Dutch speaking language area thus receive a unified account in this chapter. The presence or absence of doubling and the variation in doubling patterns is reduced to the availability of multiple copy spell out (morphology/PF), the availability of subextraction or pied piping (syntax), and the choice for specific *d*- and/or *w*-pronouns (PF/lexicalization). I furthermore argued that this analysis can successfully be extended to account for doubling involving prepositional phrases and complex *wh*-phrases, in part by adopting the notion of feature *concord*: the subextracted operator may share some features with the A-bar pronoun, as a result of which it may surface as a form different from the default form *wat* 'what'.

Notice that according to the analysis of doubling presented in this chapter, not all variation can be reduced to the lexicon or PF, but a subpart of the variation must be dealt with in syntax (*pace* Chomsky 1995, cf. section 1.2.2), namely the variation caused by the presence or absence of subextraction (or pied piping). As suggested by Marcel den Dikken, one might alternatively assume that there is no subextraction in syntax, and that the effects of subextraction (or pied piping) – i.e. the patterns of non-identical doubling – can be reduced to PF. In the remainder of this section I briefly evaluate the predictions that such a proposal makes. Consider therefore first non-identical doubling in a *wh*-Q as illustrated in (156).

(156) % **Wat** denk je **wie** het gedaan heeft?
what think you who it done has
'Who do you think has done it?' = (13b)

Instead of accounting for this sentence in terms of subextraction of the operator and double spell out (cf. *supra*), suppose that the whole DP pronoun moves successive-cyclically to the higher SpecCP and that (for some reason) the operator in SpecDP of the copy of the pronoun in the highest SpecCP is spelled out and the full copy of the pronoun in the lower SpecCP is spelled out (for recoverability reasons). This multiple spell out is allowed by the LCA (cf. section 2.2.3) as the operator higher up does not c-command the pronoun lower down, so linearization of the two copies is without problems. (Notice that if the whole DP pronoun is spelled out in the highest SpecCP, it is impossible to spell out (part of) the copy of the pronoun in the lower SpecCP, because that would constitute a violation of the LCA: the highest copy of the pronoun c-commands all its lower copies, as a result of which linearization of a structure containing two copies is impossible.) Such an analysis thus derives the same doubling patterns regarding pronoun doubling as the subextraction and double spell out analysis proposed in this chapter. Now let us consider non-identical doubling involving complex *wh*-phrases, as illustrated for a *wh*-Q in (157).

(157) % **Wat** denk je **welke man** ik gisteren gezien heb?
what think you which man I yesterday seen have
'Which man do you think I have seen yesterday?' = (44a)

If complex *wh*-phrases do not move (cf. van Craenenbroeck 2004, 2010 and see section 2.6.3), and if there is no such thing as subextraction in syntax, constructions like the one in (157) would fall outside the scope of this chapter in the sense that the boldfaced elements are not part of the same movement chain. Constructions like (157) would thus need a different analysis from constructions like (156). I think this is an undesirable result, as it would complicate the analysis of doubling, and there does not seem to be an independent reason to analyze the constructions in (156) and (157) separately.

Let us assume for the sake of argument that complex *wh*-phrases do in fact move successive-cyclically to the higher left periphery. Although the construction in (157) could then be accounted for by assuming that only the operator is spelled out in the highest SpecCP and the whole complex *wh*-phrase is spelled out in the lower SpecCP, this account of non-identical doubling runs the risk of overgeneration. There is no principled reason why only the operator can be spelled out in the higher SpecCP. More specifically, it is unclear why larger phrases like the pronoun *welke* 'which' or even the lexical NP *man* 'man' cannot be spelled out in the higher SpecCP domain, giving rise to ungrammatical chains like $[_{DP}$ *welke*$]$-$[_{DP}$ *welke man*$]$ or $[_{DP}$ *man*$]$-$[_{DP}$ *welke man*$]$.[115] An analysis that assumes doubling to be the result of spelling out a subpart of the higher copy (and that assumes that complex *wh*-phrases do in fact move

[115]In *wh*-Qs one could invoke the *wh*-requirement on the introduction of *wh*-Qs to explain the impossibility of chains like $[_{DP}$ *man*$]$-$[_{DP}$ *welke man*$]$. However, in any other movement chain that does not have a *wh*-requirement on the form of the head of the chain, it is unclear why multiple spell out of lexical NPs in excluded.

successive-cyclically to the higher left periphery) thus needs to be restricted such as to exclude multiple spell out of lexical NPs, whereas the ban on doubling with lexical NPs follows straightforwardly from a subextraction proposal.

So, at this point I do not think that an analysis of doubling in terms of full copying and spelling out a subpart of the higher copy can account for the non-identical doubling data in such a straightforward way as the subextraction and double spell out approach can. At least for now the conclusion must thus be that not all variation regarding doubling in long-distance A-bar dependencies – namely variation caused by the presence or absence of subextaction (or pied piping) – can be reduced to the lexicon or PF. This is in contradiction to the minimalist assumption that the lexicon or PF is the locus of all microvariation (cf. Chomsky 1995 a.o.). Finally, let me point out that if concord (cf. section 2.7.1) is not a surface phenomenon but an agreement (feature sharing) operation that takes place in syntax, we have yet another source of microvariation that is not located in the lexicon or PF – thereby casting even more doubt on the claim that syntactic principles are constant across (closely related) language varieties.

Appendix A: alternative analyses of doubling

In this appendix, I discuss two alternative analyses of doubling in long-distance A-bar dependencies. Whereas my analysis takes the two instances of the pronoun in doubling constructions to be members of a single chain (*single chain analysis*), these alternative analyses assume that the two instances of the pronoun are not members of the same chain, but each head their own chain (*multiple chain analyses*). I first discuss a proposal by Koopman and Sportiche (2008) that deals with special *qui* in French long-distance RCs, and that also explicitly focuses on doubling patterns in dialectal Dutch long-distance RCs. Then I evaluate a recent proposal by den Dikken (2009b) that more generally claims there is a ban on successive-cyclic *wh*-movement via SpecCP. Although this analysis does not explicitly deal with Dutch, it can be extended to cover the Dutch facts (cf. also den Dikken and Bennis 2009). I will furthermore focus on what the analysis by den Dikken (2009b) and other multiple chain analyses (e.g. Koster 2009) consider to be the nature of the embedded clause in long-distance A-bar dependencies.

To avoid confusion, a short note on terminology is in order. Whereas I make a distinction between *single* and *multiple* chain analyses, in the literature the more common division is the one between *direct dependency* and *indirect dependency* approaches. Since these approaches most prominently focus on *wh-scope marking* in long-distance *wh*-Qs – i.e. non-identical doubling involving a '*wh*-scope marker' (*wat* or its equivalent in other languages) in the higher clause and a *wh*-phrase in the lower clause – and not so much on identical doubling or on long-distance RCs, I deliberately choose not to use these terms. I will thus use the more theory neutral terms *single chain* and *multiple chain* analyses to cover all cases of doubling, and not only *wh*-scope marking.[116]

[116] As for *wh*-scope marking, the direct dependency approach (cf. van Riemsdijk 1983, McDaniel 1989, Cheng 2000 amongst many others) assumes that the scope marker and the *wh*-phrase lower down create a chain at some level of syntactic representation, by either *coindexation* or *spell out*, i.e. the *wh*-scope marker is the spell out of part of the *wh*-phrase. The direct dependency approach thus takes *wh*-scope marking constructions to be a surface alternative to long-distance *wh*-movement. The single chain analysis I assume for doubling in Dutch long-distance A-bar dependencies belongs to the family of direct dependency approaches. Indirect dependency approaches (cf. Dayal 1994, 2000 and many others) on the other hand, assume that the scope marker and the *wh*-phrase lower down are not part of the same chain. The whole embedded *wh*-clause functions as the restrictor of the *wh*-scope marker. There are many different analyses with respect to the nature of the (syntactic) relation between the scope-marker and the *wh*-phrase, among which (i) the scope-marker is base-generated as a true argument of the verb in the higher clause (e.g. Felser 2001), or (ii) the scope-marker is an expletive that is replaced by the embedded CP at LF (e.g. Fanselow and Mahajan 2000, Stepanov and Stateva 2006). The multiple chain analyses I will be discussing in this section thus belong to the family of indirect dependency approaches. See the volume by Lutz et al. (2000) for an overview of analyses of *wh*-scope marking, and see Barbiers et al. (2009) for a critical evaluation of *indirect* dependency approaches in the light of the Dutch doubling data in long-distance root *wh*-Qs.

Koopman and Sportiche (2008)

An alternative to the successive-cyclic *wh*-movement analysis of long-distance relativization has recently been proposed by Koopman and Sportiche (2008) (henceforth K&S). Although this theory seems to adequately account for the French *que*/*qui* alternation as illustrated in (158)-(159), as it stands, it cannot straightforwardly be extended to Dutch. That is, it is unclear if K&S's theory can account for all the Dutch data, because it is primarily based on French RCs and pseudo relative small clauses (PRSCs), whereas Dutch does not have PRSCs (of the same type) as in French. Moreover, K&S's account of special *qui*/*die* contexts seems to make some wrong empirical predictions with respect to the possible patterns of subject/object asymmetries in Dutch long-distance RCs (see also section 4.5 for more details).

(158) a. l'homme ***que**/**qui** _ viendra
the man that/who will come
'the man who will come' subject extraction

b. l'homme **que**/***qui** j'aime _
the man that/who I love
'the man who I love' object extraction

(159) a. l'homme **que** tu penses ***que**/**qui** _ viendra
the man that you think that/who will come
'the man who you think will come' subject extraction

b. l'homme **que** tu penses **que**/***qui** j'aime _
the man that you think that/who I love
'the man who you think I love' object extraction

On the basis of the observation that the properties of French contexts that allow special *qui* ('special contexts') are different from the properties of (French) structures that are traditionally analyzed as involving successive-cyclic *wh*-movement ('bridge contexts'), K&S argue that French special contexts should be set aside from bridge contexts involving long-distance *wh*-movement. Rather, they show that long-distance special contexts share properties with PRSCs, and in fact should be analyzed on a par with this construction. Examples of PRSCs are given in (160), in which the subject of the PRSC is italicized.

(160) a. J'ai entendu [$_{PRSC}$ *Jean* qui se faisait chahuter]
I have heard John who SE made tease
'I heard John getting teased'

b. Julie a rencontré [$_{PRSC}$ *Hélène* qui se promenait]
Julie has met Helen who walked
'Julie met Helen who was taking a walk' [K&S 2008:10]

K&S argue that extraction of *wh*-subjects out of tensed CPs is impossible in French (cf. Rizzi 1982 for Italian, and den Dikken 2009a,b for a more general

ban on extraction from SpecCP). French subject *wh*-phrases that appear to have been *wh*-moved out of their CPs are actually *wh*-moved subjects from PRSCs that are headed by *qui*: [WH [$_{PRSC}$ *qui* ...]. More specifically, what seems to be a long-distance extracted subject is actually the extracted subject of a PRSC complement of a verb belonging to a particular subset of predicates (henceforth PRED). PREDs in special contexts are very similar to French EECM verbs (Exceptional ECM verbs), which are "ECM verbs in the ordinary sense but where the Exceptional Case Marking option is limited to situations in which the ECM marked DP is *wh*-moved" (K&S 2008:9). Put differently, with EECM verbs case becomes available after *wh*-movement, as illustrated in (161).

(161) a. * on croit cet homme être malade
we believe this man to be sick

b. l'homme qu'on croit **t** être malade
the man that we believe to be sick [K&S 2008:8]

The predicate in special contexts must thus belong to a particular group of predicates: roughly predicates of saying or attitude predicates (see K&S 2008:4ff for details). The nature of the embedding predicate thus influences the appearance of special *qui*. This is illustrated for a non EECM bridge verb in (162): special *qui* is excluded in this context.

(162) a. la fille qu'il est important que tu voies
the girl that it is important that you see object extraction

b. * la fille qu'il est important **qui** vienne
the girl that it is important *qui* come subject extraction
[K&S 2008:8]

The structure in (163) illustrates that the *wh*-subject has *wh*-moved from the subject position of the PRSC to the matrix clause (instead of being moved out of the embedded CP by *wh*-movement). Case on WH$_K$ comes from the EECM verb, as a result of which it is accusative instead of nominative. This accounts for the fact that in long-distance RCs the higher clause is always introduced by *que* and never by *qui* (cf. (159)) – *que* being accusative and *qui* being nominative.

(163) WH$_K$ PRED/EECM [$_{PRSC}$ SUB$_K$ [$_{CP-REL}$ *qui* ...]]

Abstracting away from the fact that Dutch does not have PRSCs (of the same type as in French), K&S argue that since special *die* in Dutch is also found in long object relatives (in contrast to what I assume, cf. section 2.3.1), all sentences involving special *die* in Dutch should be analyzed as a kind of PRSC structure as well. This is exemplified in an abstract manner in table 2.13 – d_{ACC} stand for the accusative and d_{NOM} stands for the nominative *d*-pronoun in a particular variety of Dutch.

Table 2.13: Koopman and Sportiche (2008) – predictions for Dutch

		highest CP	special context	SUB	predicate	
subject	NP	$\boldsymbol{d}2_{ACC}$	... V	$<d2>$	$<\boldsymbol{d}1_{NOM}>$	$<d1_{NOM}>$
object	NP	$\boldsymbol{d}2_{ACC}$	... V	$<d2>$	$<\boldsymbol{d}1_{ACC}>$	.. $<d1_{ACC}>$

Notice that this analysis of subject/object asymmetries in (long-distance) RCs thus assumes a multiple chain analysis (it is compatible with an *indirect dependency approach*). It is incompatible with a single chain analysis and multiple copy spell out. To illustrate this, table 2.14 shows that the traditional successive-cyclic *wh*-movement analysis makes different predictions about the form of the *d*-pronouns involved than the PRSC analysis.[117]

Table 2.14: Successive-cylic *wh*-movement – predictions for Dutch

		highest CP	bridge context	embedded CP	extraction site
subject	NP	$\boldsymbol{d}_{NOM}$	... V	$<d_{NOM}>$ ***da(t)***	$<d_{NOM}>$
object	NP	$\boldsymbol{d}_{ACC}$	... V	$<d_{ACC}>$ ***da(t)***	.. $<d_{ACC}>$

Table 2.13 shows that the higher *d*-form is always accusative, independently of the *function* of the extracted element (subject or object), as it is always the subject of the PRSC that is extracted. The analysis of K&S thus predicts that only varieties of Dutch that make use of the relative pronoun *die* for both nominative and accusative elements can show doubling with *die* in long-distance RCs (*die-die*). Languages that make use of *die* only with nominative elements (and *dat* with accusatives), should show the pattern *dat-die*. However, I have shown elsewhere (Boef 2008a,b, Boef 2012b) that there do in fact exist varieties of Dutch in which *die* is used for short subject *and* object relativization, whereas long relativization has the form *dat-die*; see also section 4.5 for details (system VI represents the relevant pattern).[118] Such a system should not exist according to K&S's acount of special *die*/*qui* contexts (under the assumption that *dat* in these structures is indeed a *d*-pronoun). K&S's analysis of special *die* contexts and doubling thus seems to make the wrong empirical predictions.

For French, K&S present a comprehensive list of properties of special contexts, on the basis of which they show that special contexts must be different from 'standard' long-distance *wh*-configurations. Among these properties are the following: (i) the predicate of the matrix clause must belong to a particular

[117] In section 4.5, I will argue in favor of a long *wh*-movement analysis of special *die* contexts.

[118] It should be noted that not all dialects that make use of this particular pattern of long-distance relativization, only use *die* for short object relativization. Sometimes, also *dat* can be used, in which case K&S's account makes the correct prediction. Further research into the exact patterns that varieties of Dutch display in the context of long-distance RCs is necessary to settle this issue.

subset of predicates (cf. *supra*), and (ii) (negative) quantifiers and sentential negation intervening between the *wh*-operator and the instance of special *qui* in the most deeply embedded clause causes a severe degradation in acceptability. Interestingly, K&S show that both properties also seem to hold for West-Flemish (more specifically, the dialect of Lapscheure, cf. Haegeman 1992). This is illustrated in the following sentences. The West-Flemish sentences in (164) show that intervention of the negative quantifier *niemand* 'nobody' degrades the acceptability of long subject relatives that involve special *die* (164a), but is perfectly fine with long object relatives that do not involve special *die* (164b).

(164) a. ??* de studenten dat er niemand zeid *die*-n do geweest oan
the students that there nobody said *die*-PL there been were
'the students who nobody said had been there'

b. de studenten dat er niemand zeid *da*-j moet contacteren
the students that there nobody said that-you must contact
'the students who nobody said you must contact'

[West-Flemish, K&S 2008:27]

In varieties of Dutch in which special *die* shows up both in subject and in object extractions, the negative intervention effect is predicted to be found for both subject and object extraction. This prediction is borne out, as illustrated in (165).

(165) a. * de studenten die niemand geloofde *die* het verhaal verteld hadden
the students RP nobody believed *die* the story told had
INTENDED: 'the students who nobody believed had told the story'

b. * de studenten die niemand geloofde *die* ik gezien heb
the students RP nobody believed *die* I seen have
INTENDED: 'the students who nobody believed I have seen'

[Nijmegen Dutch, K&S 2008:27,28]

With respect to the sensitivity to the type of predicate in the matrix clause, (166) shows for West-Flemish and (167) shows for Nijmegen Dutch that special *die* is excluded in the context of a *desiderative* verb.

(166) a. ??* de studenten da-j zou willen *die*-n de secretaresse ipbellen
the students that-you would want *die*-PL the secretary upcall.PL
'the students who you would want to call the secretary'

b. de studenten da-j zou willen *da* de secretaresse ipbelt
the students that-you would want that the secretary upcalls
'the students who you would want the secretary to call'

[West-Flemish, K&S 2008:28]

(167) a. * de studenten die je wou *die* d'r gingen opbellen
the students RP you wanted *die* her go.PL upcall.INF
INTENDED: 'the students who you wanted to call her'

b. * de studenten die je wou *die* ik ging opbellen
the students RP you wanted *die* I go upcall.INF
INTENDED: 'the students who you wanted me to call'

[Nijmegen Dutch, K&S 2008:29]

Crucially, however, in West-Flemish the sensitivity to an intervening negative quantifier and the sensitivity to the choice of predicate seem to be independent from the presence of special *die*. Long-distance subject extraction with special *die* (as in (164a) and (166a)) seems to have more or less the same status as long-distance subject extraction without special *die*,[119] as illustrated in (168): it is severely degraded with respect to long-distance object extraction.

(168) a. ??? de studenten dat er niemand zeid *da*-n do geweest oan
the students that there nobody said that-PL there been were
'the students who nobody said had been there'

b. ??? de studenten da-j zou willen *da*-n de secretaresse ipbellen
the students that-you would want that-PL the secretary upcall.PL
'the students who you would want to call the secretary'
[West-Flemish, K&S 2008:27]

The same seems to hold for Standard Dutch, that clearly does not feature special *die*. More specifically, also Standard Dutch seems to exhibit an asymmetry between subject and object extraction – both with respect to the sensitivity to the type of embedding predicate (169) and with respect to intervening quantifiers (170).[120] Clearly, this asymmetry – that is identical to the asymmetry in West-Flemish – thus cannot be attributed to the (non-) appearance of special *die*. Consequently, the argument based on the observation that there is a difference in intervention effects between bridge contexts and special contexts loses its force.

(169) a. ?? de studenten die je wou dat de secretaresse opbellen
the students RP you wanted that the secretary upcall.PL
'the students who you would want to call the secretary'

b. de studenten die je wou dat de secretaresse opbelt
the students RP you wanted that the secretary upcalls
'the students who you would want the secretary to call'
[Standard Dutch]

[119] Special *die* is optional in long-distance subject RCs in West-Flemish, cf. Haegeman (1983) and Bennis and Haegeman (1984) for the dialect of Lapscheure (but see section 4.5).

[120] Although the Standard Dutch speakers I consulted all have a clear asymmetry between (170a) and (170b), the MPQ2-A data do not show an asymmetry between similar sentences, as illustrated in (i).

(i) a. % Dat is de man die niemand denkt dat het gedaan heeft.
that is the man RP nobody thinks that it done has
'That is the man who nobody thinks has done it.' [38/255=15%]

b. % Dat is de man die niemand denkt dat ze geroepen hebben.
that is the man RP nobody thinks that they called have
'That is the man who nobody thinks they have called.' [40/255=16%]

(170) a. ?? de studenten die niemand zei dat daar waren
the students RP nobody said that there were
'the students who nobody said were there'

b. de studenten die niemand zei dat je moet bellen
the students RP nobody said that you must call
'the students who nobody said you must call' [Standard Dutch]

It remains unclear what is responsible for this asymmetry, but it is evident that it has nothing to do with the presence of special *die*. Needless to say, a more systematic investigation of intervention effects in relation to the appearance of *die* in the lower clause of long-distance RCs is required to settle this issue.[121] In section 4.5, I come back to the issue of subject/object asymmetries and special *die* in particular, and different ways to license subject extraction more generally.

Den Dikken (2009b)

Den Dikken (2009a,b) argues that successive-cyclic movement via SpecCP is empirically and theoretically not required and therefore does not exist. Put differently, movement to SpecCP is always terminal. Consequently, what traditionally has been analyzed in terms of successive-cyclic movement via SpecCP needs to get an alternative account. Den Dikken shows that all long-distance A-bar dependencies can be formed by one of the strategies in (171).

(171) *Strategies for forming long A-bar dependencies* (den Dikken 2009b)

a. successive-cyclic movement via *v*P edges (à la Rackowski and Richards 2005)

b. resumptive prolepsis (cf. Salzmann 2006)

c. scope marking (with no/partial/full concord)

Especially the third strategy in the typology of long A-bar dependencies in (171) is relevant in accounting for the doubling patterns in Dutch long-distance A-bar dependencies. That is, all doubling patterns in long-distance A-bar dependencies reduce to (concordial) scope marking constructions (cf. den Dikken and Bennis 2009 for an explicit proposal for Dutch long-distance root *wh*-Qs). Therefore I will not focus in detail on the other two strategies in (171).

Den Dikken argues that *wh*-scope marking and *wh*-copying in long-distance *wh*-Qs are fundamentally the same constructions. They both share the presence of a *wh*-scope marker in the higher clause which may (as in *wh*-copying) or may not (as in *wh*-scope marking) enter into a concord relation (feature sharing) with the 'real' *wh*-constituent in the lower clause. These two options

[121] It should be noted that Koopman and Sportiche (2008) themselves do not offer an explanation for most of the properties of special contexts, namely intervention effects, mood restrictions and the observation that French pseudo RCs are restricted to subject RCs. In fact, they only offer an explanation for the predicate restriction on special contexts (cf. *supra*).

are schematically illustrated in (172) in which coindexing indicates a concord relation.

(172) a. SM ... $[_{vP}$ ~~SM~~ $[_{vP}$ *v* $[_{VP}$ V ~~SM~~ $[_{CP}$ XP C ... ~~XP~~ ...]]]]
plain *wh*-scope marking

b. SM_i ... $[_{vP}$ ~~SM~~ $[_{vP}$ *v* $[_{VP}$ V ~~SM~~ $[_{CP}$ XP_i C ... ~~XP~~ ...]]]]
concordial *wh*-scope marking

Plain scope marking involving pronouns surfaces as *wat-wie/die* in Dutch: the invariant scope marker is spelled out as *wat* in the left periphery of the higher clause, whereas the 'real' constituent is spelled out in the left periphery of the lower clause as *wie* or *die*. In copying constructions (with pronouns) on the other hand, the constituent in the lower clause may share features with the scope marker upstairs, allowing the latter to be spelled out differently from *wat*. Concord may be *partial*, which gives rise to identical and non-identical copying constructions. Elements that have the same surface form may thus be the spell out of different feature bundles, just as in my proposal outlined in this chapter. Conversely, concord may also be *full*, in which case we get the Standard Dutch pattern (pronoun higher up and complementizer lower down): because all features are shared with the scope marker, only the scope marker is spelled out, whereas the constituent in the lower clause remains silent (and a complementizer appears). The Standard Dutch pattern can in principle also be derived by successive-cyclic movement via *v*P edges (strategy (171a)).[122] This is all summarized in the table below.[123]

Table 2.15: Analysis of long *wh*-questions in Dutch (den Dikken 2009b)

pattern	successive-cyclic movement via edge of *v*P	plain scope marking	partial concordial scope marking	full concordial scope marking
wie-dat	+			+
wat-wie		+		
wat-die		+		
wie-wie			+	
wie-die			+	

The ungrammatical doubling pattern *wie-wat* is a case of adding features and is thus ruled out by the *Inclusiveness Condition* (cf. *supra*), i.e. assuming that *wie*

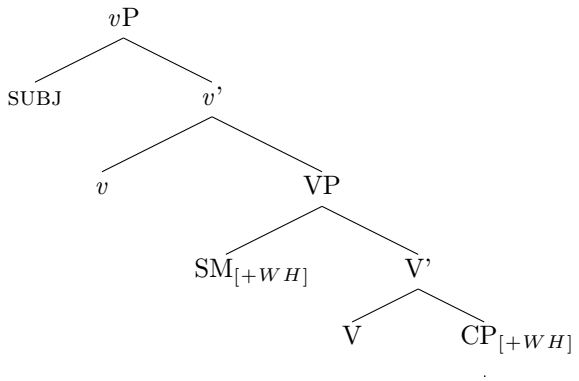
[122]Whereas in Hungarian the two strategies – full concordial scope marking and successive-cyclic movement via *v*P edges – lead to different surface forms, in Dutch the two strategies cannot be distinguished on the basis of their morphology. Which strategy is chosen is presumably dependent on economy (of computation and parsing).

[123]Notice that concord is ill-defined in the sense that it has not been worked out in terms of features, i.e. it is unclear which features the relevant elements have and which feature(s) they share under concord.

is more specified than *wat*, the latter cannot share more features with the scope marker than it has itself. The analysis of long-distance *wh*-Qs as summarized in table 2.15, can easily be extended to cover the Dutch doubling patterns in long-distance RCs – although it needs to be fleshed out what exactly it means to have a *scope marker* in a RC.

Nothing so far has been said about the nature of the embedded clause and its relation to the (scope marker in the) higher clause. Den Dikken follows a proposal made by Felser (2001), who proposes a *complex predicate approach* to *wh*-scope marking, according to which the scope marker originates in the specifier of a Larsonian V which takes the CP that contains the 'real' *wh*-constituent as its complement. According to Felser the scope marker and the CP must match with respect to the feature [+wh], i.e. there must be 'interrogative concord'. This is illustrated in (173).

(173) Felser (2001) – Complex Predicate Approach



Obviously, as also pointed out by den Dikken, the claim that the scope marker must entertain an interrogative concord relation with the CP is erroneous (for Dutch), as in long-distance *wh*-Qs (as well as in long-distance RCs) in Dutch, the lower clause may be introduced by a non-*wh* element, namely a *d*-pronoun (*die*) which cannot introduce a question in Dutch. The dependent CP (in Dutch) thus cannot be a question (*pace* Dayal 1994, Felser 2001 a.o.). This claim is further corroborated by the fact that the matrix V (*denken* 'to think' in most of the examples above) cannot take a *wh*-Q as its complement. If in *wh*-scope marking constructions the dependent CP is not a question, it cannot function as the restrictor of the scope marker, as a result of which the standard account of the semantics of *wh*-scope marking is compromised (cf. Barbiers et al. 2009:28). But if the dependent CP cannot be a question and it cannot be the declarative complement of the matrix verb (as in a direct dependency approach),[124] the natural question arises as to what is the syntactic status of

[124]If a *wh*-element introduces the lower clause, this *wh*-element clearly cannot be interpreted in that position, as the lower clause is not a question. This directly follows from my proposal as

the dependent lower CP. In the remainder of this section I will show that it is hard to give a satisfying answer to this question within an indirect dependency approach. More specifically, I will show that the dependent CP cannot be a (free) relative clause either.

It has recently been proposed that the lower clause in long-distance *wh*-Qs is a free relative clause (FRC, cf. Koster 2009 for Dutch, and Pankau 2009 for varieties of German). It is unclear what type of FRC that should be. The SAND data show that it cannot be a regular FRC in argument position (argument FRC), as there is no correlation between the form of the relative pronoun in argument FRCs and the form of the pronoun that appears in the lower clause of long *wh*-Qs (or RCs). That is to say, it is not true that an informant that uses a particular form of the pronoun in an argument FRC uses that same form in the lower clause of a doubling construction. For example, an informant may use *wie* (and disallow *die*) as a relative pronoun in an argument FRC, but allow *die* (and disallow *wie*) in the lower clause of a long-distance *wh*-Q. This pattern, as illustrated in (174), is attested in the south of North-Holland (Amsterdam and Weesp) according to the SAND1 data (Barbiers et al. 2005:90,92).

(174) a. <**Wie**/***Die**> geld heeft moet mij maar wat geven.
who/RP money has must me just what give
'Who has money, should give me some.'

b. Wie denk je **die** ik in de stad heb gezien?
who think you RP I in the city have seen
'Who do you think I have seen in the city?'
[Amsterdam/Weesp Dutch]

The lower clause in long-distance *wh*-Qs cannot be a FRC in adjunct position either, as adjunct FRCs must be introduced by a *wh*-pronoun in all varieties of Dutch (as shown by the MPQ1-A data, cf. section 4.2.2).

(175) <**Wie**/?***die**> het ook zei, hij gelooft het toch niet
who/RP it also said he believes it surely not
'Whoever said it, he doesn't believe it anyway.' [MPQ1-A data]

In addition to these mismatches in the form of pronouns in FRCs and the form of pronouns in the lower clause of long-distance A-bar dependencies involving doubling, another argument against the claim that the lower clause of long A-bar dependencies is a FRC, is the lack of *matching effects* in long-distance A-bar dependencies – a key property of FRCs (cf. Groos and van Riemsdijk 1981, van Riemsdijk 2006 a.o., and see also section 2.5.1). That is, the category of the head of a FRC must be appropriate for selection of both the matrix clause and the RC, as illustrated by the contrast in (176a) and (176b). In (176a) the PP satisfies the selectional requirements of both the matrix clause and the RC. In

outlined in previous sections, because movement of the *wh*-element to the embedded SpecCP is not assumed to take place for feature checking reasons (cf. section 2.2.2).

(176b), on the other hand, the PP that heads the RC is properly selected within that clause, but it is inappropriate as the direct object of the verb *denken* 'to think', which does not select PPs.

(176) a. Jan is verliefd [$_{PP}$ op wie Kees verliefd is].
Jan is in love on who Kees in love is
'Jan is in love with whoever Kees is in love.'

b. * Jan denkt [$_{PP}$ op wie Kees verliefd is].
Jan thinks on who Kees in love is

Now, if we were to assume that the lower clause in long-distance A-bar dependencies is a FRC, we would predict – analogously to (176b) – that (177) is ungrammatical due to the lack of a matching effect, contrary to fact (cf. section 2.7.2).

(177) % Ze vroeg **op wie** jij denkt [**op wie** hij verliefd is].
she asked on who you think on who he in love is
'She asked who you think he is in love with.'
[152/333=46%, MPQ1-B data]

Furthermore, whereas the lower clause of long-distance A-bar dependencies may be introduced by complex *wh*-phrases like *welke man* 'which man' – cf. (138) in section 2.7.1, here repeated as (178) – there are more restrictions on what may appear as the head of a FRC. More specifically, although FRCs may be introduced by complex *wh*-phrases introduced by *welke* 'which', this seems to be most natural with epithets, as evidenced by the contrast between (179a) and (179b), cf. also section 3.4.4.2.

(178) % **Wat/wie** denk je **welke man** ik gisteren gezien heb?
what/who think you which man I yesterday seen have
'Which man do you think I have seen yesterday?'
[= (138), MPQ1-B data]

(179) a. Welke onverlaat zoiets doet krijgt straf.
which miscreant such a thing does gets punishment
'Whichever miscreant does such a thing will be punished.'
[de Vries 2004:193]

b. ?? Welke man zoiets doet krijgt straf.
which man such a thing does gets punishment
'Whichever man does such a thing will be punished.'

In sum, under a multiple chain analysis of doubling in long-distance A-bar dependencies (indirect dependency approach), the lower clause cannot be the declarative complement of the matrix verb. Crucially, however, it cannot be a *wh*-Q or a FRC (argument or adjunct) either.[125]

[125] As pointed out to me by Marcel den Dikken, it is conceivable that the dependent CP is a null-headed RC of the type found in *it*-clefts (cf. den Dikken 2009c). This clause entertains

Besides it being unclear what the nature is of the lower clause in long-distance A-bar dependencies – which seems to be a more general problem that applies to any multiple chain analysis of long-distance A-bar dependencies – the analysis of den Dikken (2009b) faces another problem as well. Without additional assumptions, it is unclear how it would account for the grammaticality of doubling of a prepositional phrase that contains an A-bar pronoun, as repeated here in (180). Unless the scope marker in the higher clause can be a PP – which is unlikely, as such an assumption opens the way for the existence of all kinds of complex (*wh*)-scope markers – it is unclear how the preposition ends up in the higher clause without taking recourse to movement of the PP from the embedded clause.

(180) % **Op wie** denk je **op wie** hij verliefd is?
on who think you on who he in love is
'Who do you think he is in love with?' = (34)

---

an *asyndetic specification* relation (Koster 2000a) with the scope marker, and the operator in the RC may enter into a concord relation (mediated by the null head) with the scope marker. Although this analysis is appealing in many respects – e.g. it accounts for the fact that the RC is necessarily in radically clause-final position (like Dutch right dislocations), and for the observation that *it*-clefts and the lower clause in long A-bar dependencies share the same properties – the intended semantics of the *wh*-scope marking construction does not come about from the proposed structure. That is to say, the result of assuming that in a *wh*-scope marking construction such as (ia), the scope marker and the embedded clause are in an asyndetic specification relation is that the embedded clause does not semantically act as the *restrictor* of the scope marker, but in fact *specifies* the meaning of the scope marker, as illustrated in (ib). The only way to circumvent this problem would be to assume that there is an additional cleft structure in which the [*wh/d het is*] part is deleted at PF, as illustrated in (ic).

(i) a. wat denk je [wie het gezegd heeft]?
what think you who it said has

b. [wat: wie het gezegd heeft] denk je?
what who it said has think you

c. wat denk je [~~wie/die het is~~ wie het gezegd heeft]?
what think you who/*die* it is who it said has

This theory runs into the same problems as the theory of Koster (2009) does, the most prominent one being that the assumed ellipsis operation is not independently motivated. As noted by Lasnik (2012:3), for English there is no evidence for the claim that it has an ellipsis process by which *it is* is deleted. In fact, there seems to be suggestive evidence against this claim, as illustrated by the example in (iia). The example in (iib) shows that the same holds for Dutch.

(ii) a. [There is a knock at the door. The occupant of the room says] Who *(is it)?

b. [Er wordt op de deur geklopt. Degene in de kamer vraagt] Wie *(is het/daar)?

Furthermore, a cleft analysis seems to make the wrong empirical predictions with respect to the attested doubling patterns: the distribution of pronouns and complementizers in cleft structures is different from the distribution of pronouns and complementizers in doubling structures; see section 4.4 for RCs.

Appendix B: remaining data

In this appendix, I discuss a group of data that seem to provide evidence against the generalization that a higher copy cannot be more specified than a lower copy (cf. Barbiers 2006, Barbiers et al. 2009). These data concern doubling constructions in which the higher clause is introduced by a complex *wh*-phrase or prepositional phrase and the lower clause is introduced by an A-bar pronoun. Upon closer scrutiny, it turns out that a large part of these data – namely doubling involving complex *wh*-phrases – falls outside the scope of the analysis outlined in this chapter.

First, doubling constructions in which the higher clause is introduced by a complex *wh*-phrase whereas the lower clause is introduced by an A-bar pronoun are known to exist in child language, as first noticed by van Kampen (1997) and illustrated in (181). Closer empirical investigation (MPQ2-A) reveals that sentences like the ones in (181) are attested in colloquial Dutch as well. The sentences in (182) illustrate this for *wh*-Qs with a complex *wh*-phrase introduced by *welke* 'which'.[126]

(181) a. **Welke stift** denk je **die** ik moet nemen?
which felt pen think you RP I must take
'Which felt pen do you think I should take?'

b. **Wat voor pen** denk je **die** ik ga kopen?
what for pen think you RP I go buy
'What kind of pen do you think I will buy?'

c. **Welke jongen** denk je **die** daar loopt?
which boy think you RP there walks
'Which boy do you think walks there?'
[child Dutch, van Kampen 1997:153]

(182) a. % **Welke jongen** denk je **wat** het gedaan heeft?
which boy think you what it done has
'Which boy do you think has done it?' [54/255=21%]

b. % **Welke jongen** denk je **wie** het gedaan heeft?
which boy think you who it done has
'Which boy do you think has done it?' [112/255=44%]

c. % **Welke jongen** denk je **die** het gedaan heeft?
which boy think you RP it done has
'Which boy do you think has done it?'
[102/255=40%, MPQ2-A data]

Recall that I followed a proposal by van Craenenbroeck (2004, 2010) and assumed that complex *wh*-phrases do not act as operators and are therefore base-

[126] The equivalent of the doubling patterns in (182) is also attested in RCs, e.g. *het meisje wiens moeder jij denkt wat/wie/die het gedaan heeft* 'the girl whose mother you think what/who/RP it done has' (MPQ2-A data; between 19% and 30% of the informants accept such sentences).

generated in the left periphery. If true, the bold faced elements in the sentences in (181) and (182) can never be part of the same movement chain, as a result of which these data fall outside of the scope of the proposal made in this chapter, i.e. the proposed analysis is only concerned with doubling *within* a single movement chain. Furthermore, as too little is known about the exact properties of this construction, I leave an analysis of these data for further research. Let me only point out that the facts in (181) and (182) might be explained by assuming that the operator can be spelled out in the left periphery of the lower clause. More specifically, the complex *wh*-phrase is base-generated and spelled out higher up whereas a coindexed operator successive-cyclically moves to the higher left periphery (cf. section 3.4.1 for details). For some reason – potentially to facilitate processing – the copy of the operator in the lower SpecCP is spelled out, as abstractly illustrated in (183). The observation that the operator may be spelled out as *wie* or *die* (182b,c), in addition to the default spell out *wat* (182a), suggests again that there needs to be some sort of feature sharing (or matching) mechanism between the operator and the complex *wh*-phrase (they seem to share the feature [human] here).[127]

(183) $[_{CP}$ [pronoun NP] ~~operator$_T$~~ ... $[_{CP}$ operator$_1$ ... ~~operator$_T$~~ ...
welke man ***wat/wie/die***

Suggestive evidence in favor of the idea that a copy of an operator that successive-cyclically moved to the left periphery of the higher clause can be spelled out only in the left periphery of the lower clause, comes from long-distance left dislocation or topicalization in Dutch, as illustrated in (184). I simply assume for the moment that these constructions involve base-generation of the dislocated constituent in the left periphery – *die man* in (184) – and movement or topicalization of an operator (or 'topic pronoun', Hoekstra 1999), cf. section 3.6.7. The occurrence of the sentence in (184c) suggests that it is possible to spell out this operator/pronoun only in the left periphery of the lower clause.

(184) a. % Die man **die** denk ik dat ze geroepen hebben.
that.C man RP think I that they called have
'That man, I think they have called.' [267/333=80%]

b. % Die man **die** denk ik **die** ze geroepen hebben.
that.C man RP think I RP they called have
'That man, I think they have called.' [152/333=46%]

c. % Die man denk ik **die** ze geroepen hebben.
that.C man think I RP they called have
'That man, I think they have called.'
[178/333=53%, MPQ1-B data]

Notice that the fact that the operator may surface as *wat* in colloquial Dutch (182a) seems to provide evidence for the claim that the element that surfaces

[127]This type of concord is different from the concord briefly mentioned above, in the sense that the two elements that are sharing a feature do not originate within the same phrase.

in the lower left periphery is truly an operator and not a relative pronoun (as argued by van Kampen 2010 for child language), as *wat* cannot lexicalize a relative pronoun to a common gender human antecedent like *jongen* 'boy' independently (cf. section 2.5.2.2).

Just like doubling involving a complex *wh*-phrase, the following doubling constructions involving a prepositional phrase that contains an A-bar pronoun are attested as well (albeit rather marginally).

(185) % **Op wie** denk je **wie** hij verliefd is?
on who think you who he in love is
'Who do you think he is in love with?'
[98/380=26%, MPQ2-B data]

(186) % Dat is het meisje **op wie** je denkt **wie** hij verliefd is.
that is the girl on who you think who he in love is
'That is the girl who you think he is in love with.'
[71/380=19%, MPQ2-B data]

Because I assume that prepositional phrases containing an A-bar pronoun behave as operators and thus *move* to the left periphery (cf. section 2.7.2), the boldfaced elements in (185) and (186) are supposedly part of the same movement chain. These constructions thus seem to constitute a real counterexample to the generalization that within a single movement chain a higher copy cannot be more specified than a lower copy. Notice that these constructions most likely cannot be analyzed on a par with the constructions in (182) – i.e. base-generation of the prepositional phrase in the higher left periphery and successive-cyclic movement of a coindexed operator that is spelled out in the lower left periphery – because the argument base position is a PP whereas the operator arguably is not.

At this point I have no explanation for the grammaticality of these doubling configurations. As far as I can see, the only way of explaining their grammaticality is to invoke *remnant movement*: the DP pronoun (or the operator in SpecDP) moves out of the PP, after which the remnant PP moves further. More specifically, the DP moves to the specifier position of some functional head F^0. The DP undergoes morphological reanalysis with this head (187a), as a result of which it becomes invisible to the LCA (cf. sections 2.2.3 and 2.2.4). The chain formed by movement of the DP is not subject to deletion, as there is no visible chain to be reduced (notice that if it is the operator in SpecDP that is subextracted, the DP from which it subextracted needs to be spelled out for recoverability reasons). After movement of the DP, the remnant PP moves up (187b) and is spelled out in its final landing site, the higher SpecCP (187c).[128]

[128] Jónsson (2008) proposes a similar analysis for preposition reduplication in Icelandic (and Scandinavian more generally).

(187) a. $[_{FP}\ DP_1{+}F^0]\ [_{PP}\ P\ [DP_1]]$
(DP in SpecFP is reanalyzed with F^0)

b. $[_{PP}\ P\ [DP_1]]\ \ldots\ [_{FP}\ DP_1{+}F^0]\ [_{PP}\ P\ [DP_1]]$
(the remnant PP is fronted)

c. $[_{PP}\ P\ [DP_1]]\ \ldots\ [_{FP}\ DP_1{+}F^0]\ \sout{[_{PP}\ P\ [DP_1]]}$
(the lower copy of the PP is deleted)

Besides it being unclear why the DP (or operator in SpecDP) moves out of PP, and what the nature of F^0 is, this *remnant movement* approach faces a serious overgeneration problem. It opens the way to all sorts of subextractions and remnant movements, thereby losing a rationale for the observation that in doubling configurations the higher element is almost always less specified than the lower element, i.e. this observation becomes nothing more than a mere coincidence. Needless to say, further (empirical) research is required to fully understand the properties of these prepositional doubling constructions – most notably the properties and nature of the preposition(s) involved – and to see whether or not an analysis of these constructions along the lines of (187) is on the right track.

CHAPTER 3

Doubling and the syntax of relative clauses

3.1 Introduction

The syntax of relative clauses (RCs) has always been an important recurrent topic in generative linguistics. Almost two decades ago, the advent of Kayne's (1994) *Antisymmetry* hypothesis gave a new impulse to the research into the syntax of relativization, resulting in a debate that has been lively ever since. This chapter focuses on the syntax of *restrictive* RCs and contributes to this debate by looking at (previously undiscussed) Dutch microvariation data.[1,2]

In section 3.2, I provide a brief overview of the three main types of syntactic analyses of RCs: the head *external* analysis (HEA), the *raising* analysis (HRA), and the *matching* analysis (MA).[3] Section 3.3 shows that the Dutch doubling data – as presented in chapter 2 – pose a serious challenge for the currently most prominent analyses of RCs, namely head *internal* analyses (HIAs; *raising* or *matching*).[4] Put differently, I will show that the doubling data are most

[1] Earlier versions of parts of this chapter were presented at the Syntax Circle (Utrecht University, December 2010), ConSOLE XIX (University of Groningen, January 2011), CUNY Syntax Supper (CUNY Graduate Center, March 2011), NYU Syntax Brown Bag (March 2011), and at the workshop on 'Reconstruction effects in Relative Clauses' in Berlin (ZAS, July 2011). Parts of this chapter also appear in Boef (2012a), Boef (2012b) and Boef (to appear).

[2] I will only be concerned with the syntax of headed *restrictive* RCs and not discuss any other type of RC in detail. For an extensive overview of the different types of RCs and their properties, see e.g. de Vries (2002).

[3] It is beyond the scope of this thesis to provide an exhaustive overview and evaluation of all the proposed analyses of restrictive RCs, but see e.g. de Vries (2002), Bianchi (2002a,b), Salzmann (2006).

[4] In the context of this chapter, the term *head internal* is only used to refer to head internal

easily compatible with a HEA of RCs. In section 3.4, I present my proposal for the syntax of Dutch RCs which assumes a specific implementation of the HEA. It will be shown that this analysis is compatible with the doubling data as well as with the full range of dialectal variation in the left periphery of Dutch RCs, most prominently doubly filled COMP data.

The final part of this chapter is devoted to making a case for the HEA. In section 3.5, I show that although HIAs have gained a lot of ground in recent years, they face a variety of problems that are not encountered by a HEA. In section 3.6, I focus furthermore in detail on the most prominent argument in favor of HIAs and in disfavor of the HEA, namely the presence of *connectivity effects* between (material inside) the RC head and the RC internal gap. I show that connectivity effects cannot always be properly treated in terms of movement or c-command, i.e. *syntactic reconstruction* is not a sufficient mechanism to capture all connectivity effects, and sometimes even makes the wrong predictions. Connectivity effects thus cannot be used as a foolproof diagnostic for movement, or for HIAs of RCs more specifically. Research should therefore focus more on capturing connectivity effects without literal reconstruction, i.e. semantic accounts of connectivity. Although I will not be able to provide a definite proposal to account for connectivity effects without literal reconstruction, several (preliminary) proposals that do so are presented in the last part of this chapter.

3.2 The syntax of restrictive relative clauses: competing proposals

Restrictive RCs have been analyzed in roughly three different ways when we focus on the origin and position of the RC head: RC external, RC internal or a combination of both.[5] This section will briefly discuss these three analyses and show that besides differing in the origin of the RC head, they differ in the relation between the RC head and the RC itself (adjunction, complementation or predication), and in the way in which the (external) RC head is related to the operator or relative pronoun (predication, movement or ellipsis).

3.2.1 Head External Analysis (HEA)

The traditional approach to the syntax of RCs, the *Head External Analysis* (HEA; cf. Quine 1960, Chomsky 1977, Smits 1988, Borsley 1997 a.o.) assumes

analyses of RCs, i.e. analyses of RCs according to which the RC head is base-generated in the position of the gap inside the RC and subsequently *moves* to the left periphery. The term should not be confused with internally headed *relative clauses*, i.e. RCs in which the RC head *surfaces* in a position inside the RC (cf. footnote 41).

[5] Most analyses of RCs are variants of one of the three main types introduced here.

that RCs are CPs adjoined to (an extended projection of) the RC head noun.[6] Inside the RC, an empty operator or relative pronoun moves to the left periphery where it is linked to the external head by means of predication – indicated through coindexing in (188).[7]

(188) ... picture$_i$ [$_{CP}$ [which$_i$/Op$_i$]$_1$ he likes _$_1$]

Head External Analysis

The RC head noun is selected by an external determiner, which thus takes scope over the RC, in concordance with the scope facts in (189). In (189a) the object *two patients* allows for a narrow scope interpretation, but the object *the two patients* in (189b) – containing a definite determiner – does not. Interestingly, when the object is relativized, as in (189c), the narrow scope interpretation of *two patients* becomes available again. This indicates that the relativized element in (189c) behaves like a nominal phrase *without* a definite determiner,[8] on a par with (189a). These sentences thus suggest that the external determiner in RCs is not part of the relative clause CP itself, and accordingly scopes over the RC (the external determiner hypothesis).

(189) a. Every doctor will examine two patients. [$\forall >$*two*; *two* $>\forall$]

b. Every doctor will examine *the* two patients. [*$\forall >$*two*; *two* $>\forall$]

c. I phoned *the* two patients [that every doctor will examine _ tomorrow]. [$\forall >$*two*; *two* $>\forall$; Aoun and Li 2003:103]

Another argument in favor of the external determiner hypothesis is illustrated by the facts in (190). The sentence in (190a) shows that a definite determiner cannot normally occur in an existential *there* construction. The grammaticality of (190b) then suggests that the definite determiner cannot be part of the relativized element, because of the lack of a definiteness effect (but see the discussion in section 2.5.2.4).

(190) a. * There were *the* men in the garden.

b. *The* men that there were in the garden were all diplomats.

[Aoun and Li 2003:103]

[6] De Vries (2002:83-84) refers to this analysis as the *old standard theory*, which combines *the base-generated head hypothesis* with *the adjunction hypothesis*.

[7] In this chapter, movement dependencies are expressed by numbers, whereas coreference relations are expressed through letters (unless indicated otherwise).

[8] This does not mean that the RC internal trace (in argument position) is an NP trace. Rather, the trace inside the RC acts as a DP trace, as has been shown by Borsley (1997:632ff.) a.o. through several tests, e.g. binding (ia), control of PRO (ib), and parasitic gap licensing (ic).

(i) a. the man that t_i thought he$_i$ saw a UFO

b. the man that t_i tried PRO$_i$ to fool everybody

c. the book that Bill criticized t_i without reading pg$_i$ [Borsley 1997:632]

The traditional HEA, according to which the RC is *adjoined* to the RC head, was the standard analysis of RCs until Kayne's (1994) *Antisymmetry* hypothesis was widely adopted: the HEA is at odds with Antisymmetry because it relies on rightward adjunction – an operation that is disallowed within the Antisymmetry framework. However, not every version of the HEA assumes an adjunction structure. Besides the traditional adjunction hypothesis, the base-generated head hypothesis has also been combined with the hypothesis that the RC is a complement of N ('N-complement hypothesis', cf. Fabb 1990, Platzack 2000), with the hypothesis that the RC is a complement of D ('D-complement hypothesis', cf. Schmitt 2000, Aoun and Li 2003), and with the hypothesis that the RC combines with the RC head in a small clause structure – the RC being the predicate and the RC head being the subject (cf. den Dikken 2006a).

Notice that under the uncontroversial assumption that a noun cannot have more than one complement, the N-complement hypothesis is not very likely to be correct as the RC head can have a complement of its own: *de analyse van relatiefzinnen die Kayne heeft voorgesteld* 'the analysis of relative clauses that Kayne proposed', or *de bewering dat Jan had zitten slapen die hij probeerde te weerleggen*, 'the claim that Jan had been sleeping that he tried to refute' (Marcel den Dikken p.c.). Furthermore, also the *one* replacement test (Jackendoff 1977) – by which the proform *one* substitutes for N' (i.e. the noun and its complement), thereby distinguishing complements from adjuncts – shows that the RC cannot be a complement of the RC head. The ungrammaticality of the sentence in (191a) shows that *of chemistry* is part of N', whereas the grammaticality of (191b) shows that the RC *that love chemistry* is not part of the N', i.e. the RC is not the complement of the RC head.

(191) a. * the students of physics and the ones of chemistry
b. the students that love physics and the ones that love chemistry

Another main reason for why the HEA has been argued to have no place in current syntactic theorizing, is that it cannot straightforwardly account for connectivity/reconstruction effects. *Connectivity* (or *connectedness*) *effects* are cases in which a phrase seems to be interpreted in a position that is different from its surface position, independently of whether or not movement is involved (see e.g. Sportiche 2006). The term *reconstruction effects*, on the other hand, is strictly speaking only reserved for those connectivity effects that are found in movement constructions – *reconstruction* (originally proposed by Chomsky 1977) being an operation that places (parts of) moved material back into its position prior to movement.

To illustrate connectivity effects in RCs, consider the sentences in (192), which show idiom connectivity and variable binding connectivity respectively.

(192) a. the [**track**] that she is **keeping** _ of her expenses
b. the [picture of **his**$_i$ girlfriend] that **every man**$_i$ likes _ best
[Salzmann 2006:21-22]

Given the adjacency requirement on idiom interpretation, the RC head *track* in (192a) needs to be interpreted in the gap position inside the RC in order to be next to the verb *keeping* and give rise to the idiomatic interpretation ('keeping informed'). Similarly, for the pronoun *his* inside the RC head in (192b) to be c-commanded and bound by its antecedent *every man*, the RC head needs to be interpreted in the gap position at LF – the point at which variable binding relations are arguably established. When movement is viewed as copy and deletion (Copy Theory of Movement, Chomsky 1993 and see section 1.2.1), the connectivity effects in (192) can be easily modeled by interpreting a lower copy in a movement chain at LF (henceforth *syntactic reconstruction*, cf. Chomsky 1993, Fox 1999b a.o.). Now, given that according to the HEA of RCs, the RC head does not originate inside the RC and there is thus no copy of the RC head RC internally, connectivity effects cannot be explained by appealing to syntactic reconstruction. Rather, under a HEA of RCs, connectivity effects need to be somehow mediated by the relative operator, or they need to be accounted for by means of *semantic reconstruction* (cf. Cresti 1995, Rullmann 1995, Sharvit 1999a, Sternefeld 2001, Ruys 2011 a.o.).

Since HIAs assume that the RC head originates in a position within the RC, connectivity effects like the ones in (192) are easily captured by syntactic reconstruction. This is (somewhat simplified) illustrated in (193): activating the lower copy of the RC head at LF makes the idiomatic interpretation (193a) and the bound variable reading (193b) available.

(193) a. LF: the ~~<track>~~ that she is **keeping** <**track**> of her expenses
b. LF: the ~~<picture of his$_i$ girlfriend>~~ that **every man**$_i$ likes <picture of **his**$_i$ girlfriend> best

This straightforward treatment of connectivity effects in RCs has often been taken as one of the most important arguments in favor of HIAs, and the main reason to abandon the HEA (but see section 3.6, in which it will be shown that connectivity effects are not a foolproof diagnostic for movement and therefore cannot be used as an argument in favor of HIAs). HIAs come in two flavors: *raising* and *matching*, which will be discussed respectively in the following two subsections.

3.2.2 Head Raising Analysis (HRA)

The *head raising analysis* (HRA; cf. Schachter 1973, Vergnaud 1974, Kayne 1994, Zwart 2000, Bianchi 1999, 2000, de Vries 2002, Bhatt 2002, Henderson 2007 a.o.) assumes that the head noun is base-generated inside the RC and raises towards the matrix clause to become the RC head. The RC head is taken to be the complement of the relative pronoun or operator,[9] so something

[9]As mentioned in footnote 8, the RC internal trace acts as a DP. Therefore, most proponents of a raising analysis (or matching analysis, cf *infra*) assume that in case of zero-relativization (and *that*-relativization, under the assumption that *that* is a complementizer),

needs to happen to get the RC head in a position linearly preceding the relative pronoun/operator. This can be done by movement of the RC head internally to the relative DP, as illustrated in (194) (cf. Kayne 1994, de Vries 2002 a.o.). Alternatively, the RC head can move out of the relative DP to an extended projection of CP (e.g. Zwart 2000), or to a position outside of the relative clause CP altogether (e.g. Vergnaud 1974). Notice that these two options violate the Condition on Extraction Domain (CED; Huang 1982) or the Freezing Principle (Wexler and Culicover 1980), cf. section 3.5.3 for details.

(194) ... the $[_{CP}$ [picture$_2$ which/Op $\__2]_1$ he likes $\__1]$

Head Raising Analysis

Most implementations of the raising analysis assume the so-called *D-complement hypothesis*, according to which the RC head occupies the specifier position of the relative clause CP and this CP is a complement to D°(cf. amongst others Smith 1964, Kayne 1994, Bianchi 1999, 2000, Schmitt 2000, Zwart 2000, de Vries 2002, Aoun and Li 2003).[10] Arguments in favor of a D-complement hypothesis are mainly based on the observation that there is a close connection between the external determiner and the RC. As illustrated in (195) and (196), the external determiner concerns the definiteness of the whole structure, not only of the RC head: the determiner cannot select a particular nominal constituent (*two pictures* in (195) or *headway* in (196)), but as soon as a RC is added, the sentence becomes grammatical.[11]

(195) a. I found (*the) two pictures of John's.
b. I found the two pictures of John's *(that you lent me).

[de Vries 2002:75, based on Kayne 1994]

a DP with an *empty* D^0 moves RC internally. This empty D^0 is licensed by the external D^0 (see Bianchi 1999, 2000 a.o. for details).

[10] An analysis of RCs that assumes *raising* together with a *D-complement* structure is also referred to as a *promotion analysis* of RCs (cf. de Vries 2002 a.o.).

[11] Other cases in which the external definite determiner is ungrammatical with a simple nominal but becomes grammatical when a RC is added concern *type*-expressions (i), measure phrases (ii), resultatives (iii) and constructions with possessive *with* (iv), cf. Schmitt (2000:311-312).

(i) a. I bought <one/*the> type of bread.
b. I bought the type of bread you like.

(ii) a. Mary weighs (*the) forty-five kilos.
b. Mary weighs the forty-five kilos that Susana would love to weigh.

(iii) a. John painted the house <a/*the> nice color.
b. John painted the house the color his girlfriend likes.

(iv) a. Mary bought a house with (*the) windows.
b. Mary bought a house with the windows that she liked.

(196) a. John made (*the) headway.
b. John made the headway *(Pete made).
[based on de Vries 2002:75]

Notice that restrictive modification more generally – not only restrictive RCs – can render ungrammatical sentences like (195a) and (196a) grammatical, thereby weakening the argument that there is a close connection between the external determiner and the RC (unless all cases in (197b)-(197e) have the same underlying structure, cf. Kayne 1994). This is illustrated in (197b-e) for the sentence in (197a).

(197) a. I love (*the) Paris.
b. the Paris (that) I love
c. I love the Paris of the 19th century.
d. I love the Paris of the Impressionists.
e. I love the hidden Paris. [Alexopoulou 2006:85–86]

A similar argument in favor of a D-complement hypothesis that has been mentioned in the literature is that there seem to be selectional restrictions between the external determiner and the *type* of RC. A non-restrictive/appositive RC is incompatible with a non-specific antecedent (*any article*), and a restrictive RC is incompatible with a unique antecedent (*the queen of Holland*), as illustrated in (198).[12]

(198) a. I saw *the* queen of Holland $_{*(,)}$ who is called B.
[*restrictive/non-restrictive]
b. *Any* article $_{(*,)}$ which is about B., is interesting.
[restrictive/*non-restrictive, de Vries 2002:74]

Cross-linguistic evidence in favor of the claim that D^0 may select a finite CP, is illustrated in (199) for Polish. Although this example does not provide direct evidence in favor of the D-complement hypothesis of RC constructions,[13] it illustrates that D+CP is a possible syntactic configuration (cf. de Vries 2002:75).

(199) To, [$_{CP}$ kogo Maria widziała] jest tajemnicą.
that.NOM who.ACC Maria saw is secret
'Whom Mary saw is a secret.' [Polish, Borsley 1997:631]

[12]Selection effects between the external determiner and the RC as in (195), (196) and (198) strongly suggest that there is an interpretative dependency between the external determiner and the RC, but it is not clear *a priori* whether or not syntactic selection is involved (cf. Alexiadou et al. 2000:8).

[13]The definite article in this structure nominalizes its complex complement, and the interpretation of (199) is different from that of a RC (cf. Borsley 1997). Nevertheless, (199) illustrates that determiners may select finite clauses.

In circumnominal RC constructions, the RC head remains within the relative CP, whereas the overt determiner (if any) occurs outside of the CP. This is illustrated in (200) for Mohave. As mentioned by de Vries (2002:76), such cases seem to provide the most clear-cut evidence in favor of the idea that RCs are CPs selected by an external D°.[14]

(200) [$_{DP}$ [$_{CP}$ Hatčoq ʔavi:-m ʔ-u:ta:v]-n^{y}-č] n^{y}əʔi:l^{y}-pč.
[[dog stone.INST SUBJ.1-hit]-DEF-NOM black-REAL
'The stone with which I hit the dog was black.'
(*or* 'The dog which I hit with the stone was black.')
[Mohave, de Vries 2002:76 (from Lehmann 1984:111)]

As already mentioned in section 3.2.1, the main advantage of the HRA over the HEA is that connectivity effects can be easily accounted for by means of syntactic reconstruction: a copy of the RC head *inside* the RC is interpreted at LF. However, as noted by Munn (1994), Citko (2001), and Salzmann (2006) amongst others, (material inside) the RC head cannot always reconstruct. The grammaticality of (201) shows that the RC head cannot have been interpreted in the gap position inside the RC, as that would have led to a Principle C violation: the R-expression *Bill* would be c-commanded by the pronoun *he*.

(201) the [picture of Bill$_i$]$_1$ that he$_i$ likes $\__1$ [Munn 1994:402]

The lack of reconstruction for Principle C – more commonly referred to as the 'lack of Principle C effect' – poses a problem for the HRA, as we would expect reconstruction effects to arise across the board.

To account for the lack of Principle C effects under a *raising* analysis, Safir (1999) proposes to extend the application domain of *Vehicle Change* (VC; originally proposed by Fiengo and May 1994 for ellipsis).[15] By VC the copy of the R-expression inside the RC gets turned into a pronoun, thereby alleviating a Principle C violation. VC is assumed to apply not only in RCs, but in *all* A-bar dependencies, i.e. all A-bar dependencies are predicted to show the lack of Principle C effect (where VC applies). As illustrated in (202), for *wh*-questions (*wh*-Qs) this prediction does not appear to be borne out.

(202) * John wondered [which picture of Bill$_i$]$_1$ he$_i$ saw $\__1$.
[Munn 1994:400]

However, there exists controversy in the literature concerning Principle C connectivity effects.[16] According to Safir (1999), most *wh*-Qs indeed show a lack of Principle C effect as well, as predicted by his proposal.

[14]Notice that this only holds if a unified analysis of all types of RCs is assumed, i.e. the example in (200) suggests a D-complement hypothesis for *circumnominal* RC constructions (although (200) can be analyzed in other ways as well), but it is unclear if circumnominal RCs have the same syntax as *restrictive* RCs.

[15]See Sauerland (2003) and Salzmann (2006) for a matching analysis of RCs that employs Vehicle Change to account for the lack of Principle C effects.

[16]According to Henderson (2007), (the lack of) Principle C effects in RCs (and in constructions involving *wh*-movement more generally), do not represent a relevant diagnostic for

Besides potentially facing the problem of overgeneration (depending on the empirical basis), an account of the lack of Principle C effects under a HRA of RCs in terms of VC does not account for the lack of reconstruction effects for e.g. idiom interpretation or binding for Principle A (cf. section 3.6.2).

3.2.3 Matching Analysis (MA)

The *matching analysis* (MA; cf. Lees 1960, 1961, Chomsky 1965, Munn 1994, Citko 2001, Salzmann 2006 a.o.) combines insights from both the HEA and the HRA. In addition to a RC external head (as in the HEA), there is a RC internal representation of the RC head as well. The relative DP containing the internal RC head moves to SpecCP (as in a HRA), and under identity with the external head, PF-deletion of the noun in SpecCP is triggered, as illustrated in (203). Rejecting the N-complement hypothesis (cf. section 3.2.1 *supra*), the relative clause CP is an adjunct to the external RC head under a MA of RCs.

(203) ... the picture $[_{CP}\ [_{DP}$ which ~~picture~~$]_1\ [_{TP}$ he likes $\__1]]$
Matching Analysis

Just like the HRA, the MA accounts for connectivity effects by interpreting a RC internal copy of the RC head at LF (syntactic reconstruction). Unlike for a HRA, on the other hand, 'anti-connectivity/reconstruction' effects – like the lack of Principle C effect in (201) – are not problematic for the MA, by assuming that in some cases the RC *external* head, instead of a copy of the RC *internal* head, can be interpreted. That is to say, the lower copy of the RC internal head can exceptionally delete at LF when its content is *recoverable* from the external head (cf. Munn 1994, Citko 2001). Put differently, the lack of connectivity effects is not problematic for the MA as it can be accounted for by assuming that in RCs syntactic reconstruction is *not* the default (in contrast to *wh*-Qs, which obligatorily show reconstruction in accordance with the *Preference Principle*, Chomsky 1993). There is only syntactic reconstruction when it is forced for some reason, e.g. for idiom interpretation (192a) or variable binding (192b). That this line of reasoning is on the right track is evidenced by the examples in (204), which show that Principle C effects re-emerge when syntactic reconstruction of the RC head is forced for some other reason: for the interpretation of an idiom in (204a) and for variable binding in (204b), cf. Munn (1994), Sauerland (1998), Citko (2001) amongst others.[17]

---

movement and syntactic reconstruction, because the effects are often very weak. He argues that the absence or presence of Principle C effects in a specific construction (whether it be RCs or *wh*-Qs) should not be explained by a particular analysis of that construction, but rather by the properties of Principle C itself. Notice that although this might be true for (the lack of) Principle C effects, something additional needs to be said about the lack of reconstruction effects for e.g. idiom interpretation and Principle A (cf. section 3.6).

[17]This is an oversimplification of the facts. In a recent questionnaire study, Heycock (2011) reports that there is a dissociation of different connectivity effects, i.e. *de dicto* interpretations of adjectives, low readings of superlatives, and idiom interpretation do not cluster together

(204) a. * The [*headway* on **Mary**$_i$**'s** project]$_1$ **she**$_i$ had *made* _$_1$ pleased the boss.

b. * The [letters by **John**$_i$ to *her*$_j$]$_1$ that **he**$_i$ told *every girl*$_j$ to burn _$_1$ were published. [Sauerland 1998:71]

Before I conclude this brief overview of the three main analyses of RCs (head external, head raising, and matching) it should be mentioned that several scholars have argued for the need to distinguish different RC constructions, both within and across languages. More specifically, primarily on the basis of contradictory connectivity effects (i.e. syntactic reconstruction is not observed throughout, cf. section 3.6), it has been argued that a HEA or a matching analysis for RCs is needed whenever there are no connectivity effects, but a raising analysis is needed whenever there are connectivity effects (cf. Åfarli 1994, Aoun and Li 2003, Sauerland 1998, 2003, Hulsey and Sauerland 2006 a.o. for restrictive RCs; cf. also Carlson 1977, Heim 1987 and Grosu and Landman 1998 for the more general claim that English has both raising and matching/head external RCs). According to Åfarli (1994) for Norwegian and Aoun and Li (2003) for (some speakers of) English, the presence or absence of connectivity effects in restrictive RCs is even overtly manifested: *der*-relatives in Norwegian and *wh*-relatives in English generally do *not* show connectivity effects (head external), whereas *som*-relatives in Norwegian and *that*/ø-relatives in English generally *do* show connectivity effects (head raising).

In the next section, I will focus on doubling in Dutch long-distance RCs (cf. chapter 2) and show that these data are most easily compatible with a head *external* analysis of RCs.

3.3 Doubling: a new argument against HIAs

As extensively discussed in chapter 2, colloquial Dutch long-distance RCs may feature the presence of a doubled pronoun in the lower CP domain, i.e. the RC itself as well as the finite embedded clause may be introduced by a (non-) identical pronoun. This is illustrated for *identical* doubling with *die* in (205).

(205) % Dat is de man [$_{RC}$ **die** ik denk [**die** het gedaan heeft]].
that is the man RP I think RP it done has
'That is the man who I think has done it.'

In chapter 2, I have argued in detail that identical doubling in long-distance RCs (as well as in long-distance *wh*-Qs) is the result of successive-cyclic A-bar movement via SpecCP and multiple copy spell out. Doubling as multiple copy spell out is challenging for a head *internal* analysis of RCs (HIA; raising or matching) in the following way. A HIA predicts that the RC head that is

with binding connectivity. Apparently, connectivity effects are thus not all of a piece, i.e. connectivity does not seem to be a unitary phenomenon. In section 3.6, I will come back to different connectivity effects in more detail.

contained in the copy of the relative DP in the embedded SpecCP is overtly realized in doubling contexts, *quod non* (cf. Schippers 2006).[18,19] Put differently, whereas the relative pronoun may surface in more than one copy, the RC head in Dutch cannot surface in any but the highest copy.[20] This is illustrated in an abstract manner in (206) and (207) for a *raising* analysis à la de Vries (2002) and the *matching* analysis respectively; (208) gives the corresponding ungrammatical Dutch sentences with identical doubling of pronoun *die*.

(206) D° $[_{CP}$ [**RC head** RP]$_1$ .. $[_{CP}$ [**RC head** RP]$_1$ .. ~~[RC head RP]$_1$~~ ..
raising analysis

(207) D° **RC head** $[_{CP}$ [RP ~~RC head~~]$_1$ .. $[_{CP}$ [RP **RC head**]$_1$ .. ~~[RP RC head]$_1$~~
matching analysis

(208) a. * de $[_{CP}$ [**man die**] ik denk $[_{CP}$ [**man die**] (dat) het gedaan heeft
the man RP I think man RP that it done has
raising analysis

b. * de **man** $[_{CP}$ [**die** ~~man~~] ik denk $[_{CP}$ [**die man**] (dat) het gedaan heeft
the man RP I think RP man that it done has
matching analysis

The doubling data are especially problematic for the raising analysis of RCs, but they can in principle be accommodated by the matching analysis of RCs when the definition of *PF-deletion under identity* not only has something to say about the copy that is adjacent to the external head, but also includes the copy in the embedded SpecCP.[21] This is illustrated in (209) and (210).

[18]The underlying assumption here is that complex phrases consisting of the relative pronoun and the RC head (DP_{REL}) can move to the higher left periphery in a successive-cyclic fashion, and can undergo morphological reanalysis with the C head and be spelled out multiple times (cf. chapter 2). Recall from section 2.6.3 that doubling with complex *wh*-phrases was ruled out because complex *wh*-phrases are base-generated in the left periphery (they do not act as operators), as a result of which there are no copies of the complex *wh*-phrase that can be spelled out. Now notice that if a complex phrase consisting of an A-bar pronoun and the RC head does not act as an operator either and is base-generated in the left periphery as well, raising and matching analyses of RCs could not even exist as the (DP_{REL} that contains the) RC head is immobile, and could thus never move to the left periphery.

[19]A straightforward solution to this problem presents itself: one could assume an alternative to the standard Copy Theory of Movement, according to which lower copies in a movement chain are not full copies of the moved element but are reduced and lack internal structure (in the spirit of e.g. van Koppen 2007 or Neeleman and van de Koot 2010). Interestingly, although such a solution accounts for the fact that the RC head never surfaces in a lower copy (as it only occurs in the highest chain link), it requires a new take on connectivity effects in RCs: the interpretation of a lower copy of the RC head to account for connectivity effects in RCs (syntactic reconstruction) is not possible. Under this view of copies, a HIA thus looks very similar to a HEA (with respect to connectivity).

[20]Even though the multiple chain analyses of doubling in long-distance A-bar dependencies by Koopman and Sportiche (2008) and den Dikken (2009b) have to be abandoned given the discussion in appendix A at the end of chapter 2, it is worth mentioning here that the claim that movement to SpecCP is terminal (in RCs) is very hard to reconcile with a HIA of RCs: the RC head in long-distance RC constructions is unable to 'escape' from the lowest CP.

[21]Similarly, non-identical doubling in terms of subextraction and double spell out (cf. chapter 2) is particularly problematic for a raising analysis. The relative DP containing the

(209) D° **RC head** $[_{CP}$ [**RP** ~~RC head~~]$_1$ .. $[_{CP}$ [**RP** ~~RC head~~]$_1$ .. [~~RP RC head~~]$_1$
adjusted matching analysis

(210) de **man** $[_{CP}$ [**die** ~~man~~] ik denk $[_{CP}$ [**die** ~~man~~] (dat) het gedaan heeft
the man RP I think RP that it done has

I will not attempt to make the MA of RCs compatible with the doubling data for two reasons. First, the MA faces more problems than accounting in a straightforward way for the doubling data (cf. section 3.5). It is worth mentioning here that the *PF deletion under identity* operation is a bit suspicious in itself. The precise properties and workings of this deletion operation (by which the RC internal representation of the RC head gets deleted) remain unclear and – to the best of my knowledge – have never been fully worked out. Bhatt (2002:77-79) notes that under a MA it is unclear why the external head is pronounced and why the internal head is *obligatorily* deleted (in contrast to the *optional* nature of other kinds of ellipsis).[22] The notion *PF deletion under identity* thus needs to be properly defined.

Second, the doubling data follow from a head *external* analysis of RCs without any additional stipulations. More specifically, the problem of the impossibility of spelling out the RC head in the lower CP domain is non-existent if a HEA of RCs is assumed, as there is no copy of the RC head inside the RC in the first place. This is illustrated in (211).

(211) a. D° **RC head** $[_{CP}$ **RP**$_1$ .. $[_{CP}$ **RP**$_1$ .. **RP**$_T$ ..
b. de **man** $[_{CP}$ **die** ik denk $[_{CP}$ **die** het gedaan heeft
the man RP I think RP it done has
Head External Analysis

Having shown that pronominal doubling in long-distance RCs is most easily compatible with a HEA of RCs, the next section outlines the specific implementation of the HEA that I assume for Dutch RCs. I show that this structure is compatible with the doubling data, as well as with the full range of dialectal variation in the left periphery of RCs, namely doubly filled COMP data.

RC head and the relative pronoun moves to the lower left periphery, after which the operator subextracts and gets spelled out higher up. The RC head is thus trapped inside the lower CP and never reaches the higher CP, as abstractly illustrated in (i).

(i) D° $[_{CP}$ **operator**$_2$ .. $[_{CP}$ [**RP**$_2$ **RC head**]$_1$ ... [~~RP RC head~~]$_1$ ...

[22]Unless, as noted by Bhatt (2002:77-78), this type of deletion/ellipsis is thought of as being on a par with the type of ellipsis found in the domain of *comparative deletion*, as illustrated in (ii), cf. Kennedy (2000).

(i) a. The galaxy contains more stars than the eye can see.
b. * The galaxy contains more stars than the eye can see stars.

(ii) The galaxy contains more ***stars*** than $[_{CP}$ [~~*DP stars*~~]$_T$ the eye can see [~~*DP stars*~~]$_T$].

3.4 The syntax of Dutch relative clauses

This section presents my proposal for the syntax of restrictive RCs in Dutch. As the analysis is inspired by a proposal made by van Craenenbroeck (2004, 2010) regarding the structure of the left periphery in embedded *wh*-Qs in Dutch, I first briefly present this proposal in section 3.4.1. In section 3.4.2, I focus on the differences between *wh*-Qs and RCs, and in section 3.4.3, I present my proposal for the syntax of restrictive RCs in Dutch in detail. Section 3.4.4 provides substantial empirical support for the proposal.

3.4.1 The left periphery of Dutch embedded questions

Before I present the proposal by van Craenenbroeck (2004, 2010), a brief note on doubly filled Comp (henceforth DFC) is in order. The Doubly Filled Comp Filter (henceforth DFCF; Chomsky and Lasnik 1977), as given in (212), traditionally refers to a ban on the simultaneous spell out of a *wh*-phrase and a complementizer in the left periphery (Comp) of an embedded clause.[23] This filter can explain the fact that in (modern standard) English *wh*-Qs and RCs, the *wh*-pronoun and the declarative complementizer *that* cannot cooccur, as illustrated in (213) and (214).

(212) DFCF: *[$_{COMP}$ *wh*-phrase complementizer]
[Chomsky and Lasnik 1977:435]

(213) a. I wonder **who** has done it.
b. * I wonder **who that** has done it. [modern Standard English]

(214) a. That is the man I saw.
b. That is the man **who** I saw.
c. That is the man **that** I saw.
d. * That is the man **who that** I saw. [modern Standard English]

With the notion *DFC* I will refer to configurations in which both the head and the specifier of one and the same projection in the left periphery are spelled out, without this having an effect on the semantics of the clause, i.e. spelling out two elements in the left periphery, as opposed to spelling out only one, does not make a semantic contribution. It is well known that a lot of (non-standard) languages do not obey the DFCF, as was already noted by Chomsky and Lasnik (1977). This is also true for non-standard Dutch: in different varieties of Dutch a relative or an interrogative pronoun can be followed by one (or two) complementizer(s), as illustrated in (215) for embedded *wh*-Qs (cf.

[23] See Koopman (1996, 2000) for a *Generalized* Doubly Filled Comp Filter, which prohibits the specifier and the head of *any* projection to be filled with lexical material simultaneously. I will not discuss this proposal here, but only point out that a generalized DFCF results in an explosion of functional projections (to accommodate all overt specifiers and heads), the identification/nature of which is in need of further research.

SAND1 data, MPQ1-A data).[24] The example in (216b) shows the presence of two complementizers in embedded yes/no questions.[25]

(215) a. Ik vraag me af **wie** het gedaan heeft.
I wonder who it done has
'I wonder who has done it.'

b. % Ik vraag me af **wie of** het gedaan heeft.
I wonder who whether it done has
'I wonder who has done it.'

c. % Ik vraag me af **wie dat** het gedaan heeft.
I wonder who that it done has
'I wonder who has done it.'

d. % Ik vraag me af **wie of** **dat** het gedaan heeft.
I wonder who whether that it done has
'I wonder who has done it.'

(216) a. Ik vraag me af **of** hij dit gedaan heeft.
I wonder whether he this done has
'I wonder whether he has done this.'

b. % Ik vraag me af **of** **dat** hij dit gedaan heeft.
I wonder whether that he this done has
'I wonder whether he has done this.' [Bennis 1997:353]

Notice that (216b) is in fact not an instance of DFC in the strict sense, as we are not dealing with spelling out both the head and the specifier of a single CP projection. In the literature, there has been some debate on whether *of dat* 'whether that' in constructions like (215d) and (216b) is a morphologically complex complementizer (as argued by Sturm 1996) or whether *of* and *dat* are two separate heads, each heading their own functional projection in the left periphery (cf. Hoekstra 1993a,b, Hoekstra and Zwart 1994, Zwart and Hoekstra 1997, van Craenenbroeck 2004, 2010). I believe the latter view is correct, and I will therefore briefly mention here some of the arguments in favor of *of* and *dat* being two separate functional heads (see chapter 4 for detailed discussion on the nature of complementizer *dat*).

First, claiming that *of dat* is a variant of the interrogative complementizer *of* is not very helpful in explaining the observation that *of* and *dat* also appear as complementizers separately (Bennis 2000:45) and that each complementizer is associated with its own semantic effect (or the lack thereof), as illustrated in (217).

[24] Whereas the MPQ1-A data do not show a clear geographic distribution of the sentence in (215c), the SAND1 data (Barbiers et al. 2005:16) show that this DFC pattern, i.e. *dat* following a *wh*-pronoun, is primarily attested in the Dutch speaking part of Belgium.

[25] Throughout this thesis I gloss the element *of* as 'whether', not as 'if'. The reason for this is that although *of* shares properties with both complementizers (*whether* and *if*), unlike *if*, *of* cannot be used in conditional sentences (*als* must be used in those cases).

(217) a. het feit <*dat*/ **of*> hij op tijd is
the fact that whether he on time is
'the fact that he is on time'

b. de vraag <*of*/ **dat*> hij op tijd is
the question whether that he on time is
'the question whether he is on time' [Hoekstra 1993b:193]

Second, in some varieties of Dutch, a *wh*-phrase can split up the complementizer sequence *of dat*.

(218) Ik weet niet *of* met wie *dat* Jan oan et proate was.
I know not whether with who that Jan on it talk was
'I don't know who Jan was talking to.'
[Strijen Dutch, van Craenenbroeck 2004:35]

Third, the complementizer sequence *of dat* can be broken up by conjunction reduction, as illustrated in (219), unlike what is the case with real syntactic heads, cf. (220).

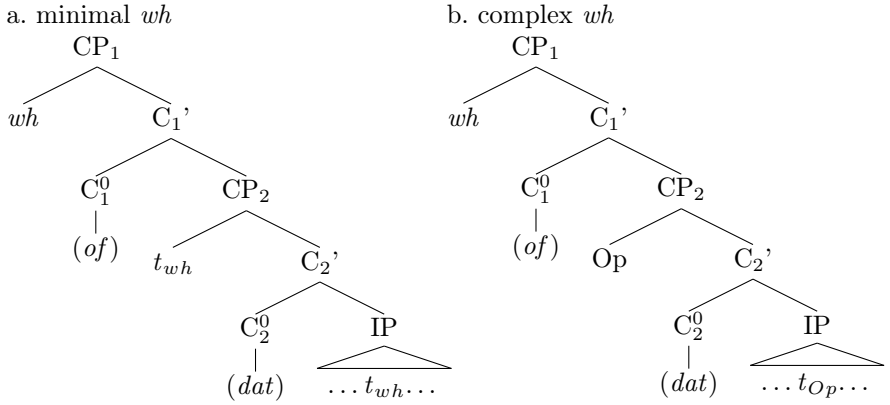
(219) Ik vraag me af [*of* [*dat* Ajax de volgende ronde haalt] en
I wonder whether that Ajax the next round reaches and
[*dat* Celtic verslagen kan worden]].
that Celtic beaten can be
'I wonder whether Ajax will reach the next round and whether Celtic can be beaten.' [colloquial Dutch, Hoekstra 1993b:193]

(220) a. * [*om*-[*dat* het nu regent] en [-*dat* het straks gaat sneeuwen]]
because it now rains and -because it later will snow
'because it rains now and it will snow later'

b. * [*in*-[*dien* hij koning is] en [-*dien* zij koningin is]]
if he king is and -if she queen is
'if he is king and she is queen' [Hoekstra 1993b:192]

Based on, but somewhat diverging from proposals by Reinhart (1981), Hoekstra (1993a,b), Hoekstra and Zwart (1994), Zwart and Hoekstra (1997), and Bennis (1997, 2000), van Craenenbroeck (2004, 2010) argues for a split CP domain (cf. Bhatt and Yoon 1992, Rizzi 1997, Baltin 2010 amongst many others) in Dutch, to, among other things, accommodate the DFC data in (215) and (216). He proposes the structures in (221).

(221) Van Craenenbroek's (2004, 2010) analysis of the left periphery in Dutch embedded (*wh*-)questions



The higher CP projection, CP_1, is optionally headed by *of* and related to *clause typing* (as in Cheng 1991). In the relevant literature this projection is also referred to as ForceP (Rizzi 1997), TypP (Bennis 1997, 2000), or MoodP (Bhatt and Yoon 1992). The lower CP projection, CP_2, can optionally be headed by *dat*, and it is the projection in which operator-variable dependencies are created (if any). This lower CP projection has also been argued to be the projection in which *subordination* is expressed (SubP: Bennis 1997, 2000, and cf. Boef to appear).

As already mentioned in section 2.6.3, van Craenenbroeck (2004, 2010) attributes the well known syntactic differences between minimal *wh*-phrases (bare *wh*-pronouns and PPs containing them) and complex *wh*-phrases to their operator/non-operator status (instead of D-linking, as Pesetsky 1987). Minimal *wh*-phrases act as operators, whereas complex *wh*-phrases do not (see section 2.6.3 for arguments in favor of this claim). As a result of this different operator status, *wh*-pronouns move through $SpecCP_2$ to $SpecCP_1$, whereas *wh*-phrases are base-generated in $SpecCP_1$ while a coindexed empty operator moves to $SpecCP_2$. This is illustrated in (221a) and (221b) respectively.

Van Craenenbroeck's (2004, 2010) analysis of the left periphery in Dutch embedded questions predicts the existence of the DFC patterns as given in table 3.1. As the rightmost column of the table shows, the predictions are borne out: all and only the patterns in table 3.1 are attested. That is to say, the analysis correctly rules out (i) the order *dat of*, (ii) constructions in which a simplex or complex *wh*-phrase follows the two complementizers, and (iii) constructions in which a complex *wh*-phrase surfaces in a position between *of* and *dat*, i.e. whereas in Strijen Dutch it is possible to split the complementizer sequence *of dat* with a simplex *wh*-phrase, as was illustrated in (218), it is impossible to do so with a complex *wh*-phrase. Why exactly a simplex *wh*-phrase in Strijen Dutch may stay in $SpecCP_2$ is subject to further research.

(222) * Ik vroag me af ***of*** welke jonge ***dat*** die maisjes gistere
I ask me PRT whether which boy that the girls yesterday
gezien hebbe.
seen have
INTENDED: 'I wonder which boy the girls saw yesterday.'
[Strijen Dutch, van Craenenbroeck 2010:245]

Table 3.1: Predictions of van Craenenbroeck's (2004, 2010) analysis

$SpecCP_1$	C^0_1	$SpecCP_2$	C^0_2	occurrence
	of			Standard Dutch
	of		*dat*	colloquial Dutch (Bennis 1997:353)
wie		t_{wie}		Standard Dutch
wie	*of*	t_{wie}		colloquial Dutch (Bennis 1997:353)
wie	*of*	t_{wie}	*dat*	colloquial Dutch (Bennis 1997:353)
wie		t_{wie}	*dat*	southern Dutch/Belgian dialects (Haegeman 1992:55, Bennis 2000:44)
welke man		Op		Standard Dutch
welke man	*of*	Op		colloquial Dutch (van Craenenbroeck 2004:44)
welke man	*of*	Op	*dat*	colloquial Dutch (Hoekstra 1993b:59)
welke man		Op	*dat*	colloquial Dutch (Hoekstra 1993a:161)
	of	*(met) wie*		Strijen Dutch (van Craenenbroeck 2004:34)
	of	*(met) wie*	*dat*	Strijen Dutch (van Craenenbroeck 2004:35)

A natural question that arises at this point is whether van Craenenbroeck's (2004, 2010) analysis of the left periphery in embedded Qs can be extended to the left periphery in restrictive RCs. In the next section I will show that there are some striking differences between embedded *wh*-Qs and RCs in Dutch, which will form the basis for my proposal of the left periphery of Dutch RCs.

3.4.2 Embedded *wh*-questions versus relative clauses

A first difference between embedded *wh*-Qs and restrictive RCs is that complex *wh*-phrases introduced by *welke* 'which' are not allowed in restrictive RCs (223), in contrast to non-restrictive/appositive RCs (224), cf. Haeseryn et al. (1997:331), de Vries (2004).

(223) a. * Ze zag een man *welke stakkerd* zijn been had gebroken.
she saw a man which wretch his leg had broken

b. * Ze las een boek *welke roman* door Reve was geschreven.
she read a book which novel by Reve was written
[de Vries 2004:200]

(224) a. Ze zwaaide naar Joop, *welke stakkerd* zijn been had gebroken
she waved at Joop, which wretch his leg had broken
bij het skieën.
at the skiing
'She waved at Joop, which wretch has broken his leg during skiing.' [de Vries 2004:199]

b. "In de ban van de ring", *welk boek* van Tolkien zeer populair
in the spell of the ring which book by Tolkien very popular
is, is verfilmd.
is has been filmed
'"The lord of the rings", which book by Tolkien is very popular, has been filmed.' [de Vries 2004:193]

Another difference between embedded *wh*-Qs and restrictive RCs in Dutch is that the string *wie of dat* 'who whether that' is not (or very marginally) attested in restrictive RCs (cf. SAND1 data, further field research (Boef to appear), and MPQ1-A data). This is illustrated in (225d).[26] Since *d*-pronouns cannot introduce *wh*-Qs independently (cf. chapter 2) and *d*-pronouns are incompatible with complementizer *of* (cf. *infra*), I will not be concerned here with DFC involving a *d*-pronoun, but see section 3.4.4.1 for more discussion on DFC patterns in Dutch.

(225) a. % Dat is de man **wie** het verhaal verteld heeft.
that is the man who the story told has
'That is the man who told the story.'

b. % Dat is de man **wie of** het verhaal verteld heeft.
that is the man who whether the story told has
'That is the man who told the story.'

c. % Dat is de man **wie dat** het verhaal verteld heeft.
that is the man who the story told has.
'That is the man who told the story.'

d. * Dat is de man **wie of** **dat** het verhaal verteld heeft.
that is the man who whether that the story told has

[26] The sentence in (225d) was not explicitly tested in the SAND project. To accurately check the status of this construction, I did a follow up telephonic interview (cf. Boef to appear). I selected from the SAND corpus all locations that allow *wie of dat* 'who whether that' in embedded *wh*-Qs and all locations that make use of relative pronoun *wie* 'who' in short subject or short object RCs. There are only six locations that allow both *wie of dat* in *wh*-Qs and that make use of relative pronoun *wie*. The informants in all six locations strongly rejected the construction in (225d). However, the MPQ1-A data show some attestations of *wie of dat* in RCs, but since the amount of attestations is very low and they do not show a clear geographic pattern, I take these attestations to represent noise in the data set (cf. section 1.3.2). Although the DFC patterns *wie of* and *wie dat* in RCs are also attested only marginally in MPQ1-A, these DFC patterns are attested considerably more frequently than the pattern in (225d) and it is known from the literature that such constructions occur in spoken Dutch (cf. table 3.2), hence the % sign in front of the sentences in (225b) and (225c).

On the basis of these two differences, I claim that the RC head in RCs is merged in the position where complex *wh*-phrases introduced by *welke* 'which' are merged in embedded *wh*-Qs, namely SpecCP$_1$/SpecForceP (cf. Schmitt 2000, Aoun and Li 2003 for similar proposals).[27] More specifically, I propose that the RC head is base-generated in the highest SpecCP of the RC itself, whereas the relative pronoun moves to the lower SpecCP. This analysis thus combines the *base-generated head hypothesis* with the assumption that the RC head originates *inside* the RC as in a HIA. Crucially, my analysis does not assume with a HIA (raising or matching) that the RC head originates in the position of the gap inside the RC and moves to the left periphery. Due to the lack of movement of the RC head to the left periphery (raising), I will refer to my analysis as a HEA, even though the RC head is not strictly speaking base-generated in a position *outside* of the RC itself.

The proposed analysis explains that complex *wh*-phrases introduced by *welke* 'which' are absent in restrictive RCs, as such phrases are in complementary distribution with the RC head. It furthermore accounts for the lack of the string '*wh*-pronoun *of dat*' in RCs, as there is simply not enough space to fit both complementizers to the right of the relative pronoun. The next sections deal in more detail with the analysis of Dutch restrictive RCs and the predictions it makes.

[27] Rizzi (1997:289) shows (after Cinque 1979) that in Italian the relative operator seems to target a higher position in the left periphery than the interrogative operator (cf. den Dikken 2003 for Hungarian). More specifically, as illustrated by the contrasts in (i) and (ii), relative operators must precede topics, whereas interrogative operators must follow topics in *root wh*-Qs. As the judgments in (i)-(ii) do not carry over to Dutch, I will not be concerned with this difference between relative and interrogative operators here. Furthermore, it is important to point out that once embedded *wh*-Qs (instead of root *wh*-Qs) are taken into account, the situation in Italian becomes more like Dutch: interrogative operators may also *precede* topics – just like relative operators – as can be seen in (iii). This seems to indicate that in Italian *embedded* clauses the relative operator and the *wh*-operator may target the same position – just like I argue is the case for Dutch.

(i) a. Un uomo a cui, il premio Nobel, lo daranno senz'altro.
'A man to whom, the Nobel prize, they will give undoubtedly.'

b. * Un uomo, il premio Nobel, a cui lo daranno senz'altro.
'A man, the Nobel prize, to whom they will give undoubtedly.'

(ii) a. * A chi, il premio Nobel, lo daranno?
'To whom, the Nobel Prize, will they give it?'

b. Il premio Nobel, a chi lo daranno?
'The Nobel Prize, to whom will they give it?'

(iii) a. Mi domando, il premio Nobel, a chi lo potrebbero dare.
'I wonder, the Nobel Prize, to whom they could give it.'

b. ? Mi domando a chi, il premio Nobel, lo potrebbero dare.
'I wonder to whom, the Nobel Prize, they could give it.'

3.4.3 The proposal

As pointed out in section 3.3, a head *external* analysis of RCs can straightforwardly account for pronominal doubling in RCs, whereas head *internal* analyses need additional assumptions to accommodate these data. Therefore, I propose an analysis of RCs in Dutch that assumes that the RC head is base-generated 'outside' the RC. Instead of assuming an adjunction structure as in the traditional HEA (cf. section 3.2.1), I will follow common practice and assume that the relative clause CP is the complement of the external determiner (*D-complement hypothesis*, cf. section 3.2.2). On the basis of the two differences between embedded *wh*-Qs and RCs as outlined in the previous section, I will furthermore assume that the RC head is base-generated in the higher SpecCP (cf. Aoun and Li 2003 a.o.), whereas a relative pronoun or operator moves to the lower SpecCP (notice that a split CP structure is forced by the D-complement hypothesis together with a HEA of RCs in order to provide enough space for both the RC head and the relative pronoun or operator).[28] This is illustrated in (226).[29]

[28]It is unclear what exactly is the nature of the empty operator in Dutch RCs (in varieties that make use of an empty operator – mostly southern Dutch, cf. SAND1 data: Barbiers et al. 2005, and see section 4.5). It most likely is not a true empty operator. As argued by den Dikken (1995), an indirect object cannot undergo true empty operator movement. However, varieties of Dutch that make use of an empty operator in RCs do in fact allow indirect objects to undergo empty operator movement, as illustrated in (i) for West-Flemish.

(i) den vent [*Op* da Jan peinst [t_{Op} da-n ze t_{Op} dienen boek beloofd een]]
the man that Jan thinks that-3PL they that book promised have
'the man to whom Jan thinks they have promised that book'
[West-Flemish; Liliane Haegeman p.c.]

I have not been able to come up with an independent, foolproof diagnostic (i.e. something different from connectivity effects, cf. *infra*) for determining the nature of the empty operator: either a covert counterpart of the *pronoun* (as in a HEA) or a covert counterpart of the *RC head* (as in a MA).

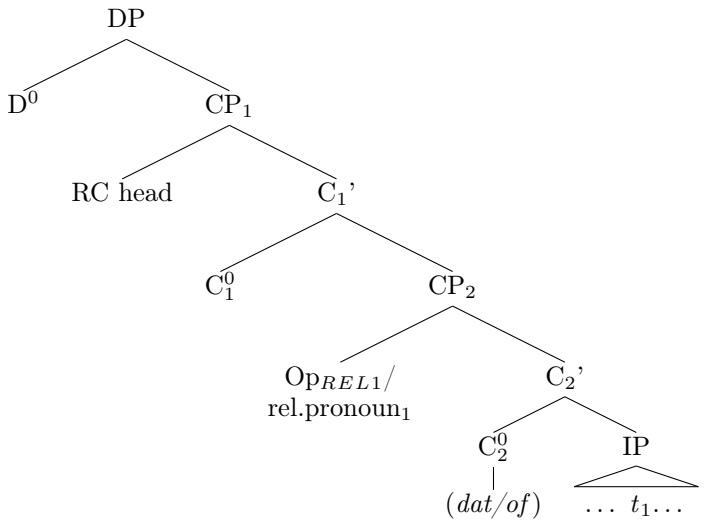
[29]The analysis proposed in (226) shares several features with the analysis of restrictive RCs (in Mandarin Chinese) as proposed by den Dikken (2006a:chapter 5), namely a D-complement structure, the lack of movement/raising of the RC head, and the assumption that the RC is linked to the RC head through predication. More specifically, according to den Dikken's analysis D° takes a small clause complement of which the RC head is the subject and the RC itself is the predicate. This is illustrated in (i). The functional head of the small clause (the RELATOR in den Dikken's terminology) may be realized in languages like Mandarin Chinese, as illustrated in (ii).

(i) D° [$_{SC}$ [RC head] [RELATOR [$_{CP}$ RC]]] [adapted from den Dikken 2006a:424]

(ii) wo mai de shu
I buy DE book
'the book that I bought' [Mandarin Chinese; den Dikken 2006a:240]

Notice that the observation that there are languages in which the RC head and the RC itself may be separated by intervening material (the functional element *de* in the case of Mandarin Chinese) is highly problematic for a *raising* analysis of RCs à la Kayne (1994) or de Vries (2002), because the RC head in SpecDP_{REL} in SpecCP will always immediately precede the relative pronoun/operator (D_{REL}) that introduces the RC. The structure I propose in (226)

(226) The structure of restrictive RCs in Dutch
cf. Schmitt (2000), Aoun and Li (2003)



Just as in embedded *wh*-Qs, the head of the higher CP_1 in (226) (or ForceP, as in Aoun and Li 2003) is related to *clause typing* in the sense that only elements that are specific for RCs are allowed to occur there. The lower CP_2 is the layer in which *operator-variable dependencies* are created. As there is no designated element (C_{REL}) in Dutch that is responsible for clause typing in RCs (cf. Wiltschko 1998), C^0_1 is always empty in Dutch. More generally, Dutch does not have overt clause typers in other constructions either, e.g. root *wh*-Qs, topic/focus constructions; in those constructions the verb seems to move to the C^0_1 position.[30] So, the relative pronoun or complementizer (zero-relativization is not allowed in Dutch, cf. Dekkers 1999, SAND1 data: Barbiers et al. 2005) always directly follows the RC head in Dutch.[31]

Movement of the relative operator to $SpecCP_2$ turns the proposition into a predicate (*lambda/predicate abstraction*). The RC is related to the RC head by means of *Predicate Modification* (Heim and Kratzer 1998:95), which semantically amounts to *set intersection*. More precisely, the RC denotes a set which needs to be intersected with the set denoted by the RC head to get the right interpretation.[32] Given that the prerequisite for the application of set inter-

is capable of accommodating data like (ii): the DE element may target the C^0_1 position in between the RC head and the RC itself.

[30]This observation raises an interesting research question for further cross-linguistic study: do languages that have an overt clause typer in RCs, also have overt clause typers in other configurations, and vice versa?

[31]The only variety of Dutch that – to the best of my knowledge – does not adhere to this generalization is Amsterdam Dutch as reported by Hoekstra (1994); see section 3.4.4.3.

[32]As pointed out to me by Radek Šimík, one could alternatively take C^0_1 to be a *function* that takes two predicates as its arguments: the RC head and the RC (after which the relative

section/predicate modification is that the RC and the head of the RC be two independent constituents, set intersection can successfully apply in the structure in (226).[33,34]

3.4.4 Empirical support

Even though my proposal is related to several well argumented existing accounts of restrictive RCs (e.g. Schmitt 2000, Aoun and Li 2003), it is in need of (further) empirical support. The following sections provide several arguments in favor of the proposal outlined above.[35]

3.4.4.1 Doubly filled Comp

Besides being compatible with the pronominal doubling data from chapter 2 (cf. section 3.3), the structure in (226) accounts for the whole range of dialectal

CP combines with the external determiner). This function would be of the following type: $\lambda P \lambda Q \lambda x[P(x) \wedge Q(x)]$. Under such an account, the RC head thus combines with the RC by *Function Application*.

[33]Zwart (2000:378) notes that Kayne (1994) or de Vries (2002) style analyses of RCs – according to which the RC head does not move out of the DP_{REL} – do not meet this prerequisite, as the RC is not a constituent *to the exclusion of* the RC head. Raising analyses of this type thus run into problems of interpretation: it is not possible to derive the interpretation of the RC construction by means of set intersection of two sets.

[34]Although Zwart (2000) assumes a *raising* analysis for Dutch restrictive RCs, his analysis is very similar in spirit to the one proposed here. Particularly, Zwart assumes a split CP structure for Dutch that has three CP layers: (i) the highest CP_1, targeted by the RC head and optionally headed by *als* 'if/when', (ii) the intermediate CP_2, targeted by *wh*-pronouns and optionally headed by *of* 'whether', and (iii) the lowest CP_3, targeted by *d*-pronouns and optionally headed by *dat* 'that'. Movement of the RC head out of the DP_{REL} (in $SpecCP_2$ or $SpecCP_3$) into $SpecCP_1$ is triggered by the need to establish a configuration in which the RC head and the RC can be two separate constituents, so set intersection can properly apply. Notice that besides running into all the problems any *raising* analysis of RCs runs into (cf. section 3.5), the structure in (i) overgenerates: there is no (structural) explanation for why the string *wh of dat* 'wh whether that' does not occur in RCs (whereas it does in *wh*-Qs, cf. section 3.4.2).

(i) [$_{CP1}$ **RC head** (*als*) [$_{CP2}$ *wh*-pronoun (*of*) [$_{CP3}$ *d*-pronoun (*dat*)]]]

[35]I have argued elsewhere (Boef to appear) that the left periphery of RCs is reduced with respect to the left periphery of embedded (*wh*)-questions, i.e. the higher CP layer – the layer that is related to clause typing and that hosts complex *wh*-phrases introduced by *welke* 'which' – is missing in restrictive RCs (cf. Reinhart 1981 for a proposal very similar in spirit). One reason to think that this claim is on the right track is that there is arguably no clause typing in Dutch restrictive RCs. First, RCs in Dutch are nothing more than declarative subordinate clauses that contain a variable, so it is hard to imagine how the clause should be typed exactly. Second, if there were clause typing in RCs, it would probably be the relative pronoun or operator that is doing the clause typing, but there is no such thing as a relative pronoun in Dutch (cf. Wiltschko 1998 and see chapter 4 for discussion). Although this analysis is appealing at first glance, the main reason for why I abandoned it in favor of the analysis in (226) is that the latter is compatible with a D-complement hypothesis, it can accommodate the Amsterdam Dutch data (cf. section 3.4.4.3), and it makes more sense from a cross-linguistic point of view (cf. section 3.4.4.4).

variation that is found in (the left periphery of) Dutch RCs (Boef to appear). The specific patterns that the structure in (226) predicts to exist are given in the following table. For the moment, I abstract away from the potential lexicalization of C_1^0 in Dutch, as in the overwhelmingly vast majority of cases, varieties of Dutch do not spell out C_1^0 (but see section 3.4.4.3 for an exception to this generalization).

Table 3.2: Predictions of relative clause structure (226), cf. Boef (to appear)

$SpecCP_1$	C_1^0	$SpecCP_2$	C_2^0	occurrence
RC head		d_{REL}		Standard Dutch
RC head		d_{REL}	*dat*	Waasland Dutch (SAND1 data)
RC head		d_{REL}	*of*	*unattested*
RC head		wh_{REL}		colloquial Dutch (MPQ1-A data)
RC head		wh_{REL}	*dat*	southern Dutch (SAND1 data)
RC head		wh_{REL}	*of*	colloquial Dutch (Hoekstra 1993b:197)
RC head		Op_{REL}		*unattested*
RC head		Op_{REL}	*dat*	Vlaams-Brabant (SAND1 data)
RC head		Op_{REL}	*of*	*unattested*

As the rightmost column in this table shows, most predictions in table 1 are borne out, but some predicted patterns are unattested. I claim that these patterns are unattested on independent grounds. First, [Op_{REL}] in table 3.2 is unattested because in all varieties of Dutch there needs to be at least one overt element in the Comp domain of restrictive RCs (and more generally, in all embedded clauses in Dutch, cf. *supra*). Second, the patterns [Op_{REL} *of*] and [*d*-pronoun *of*] in table 3.2 are unattested, because *of* in RCs is only licensed by an overt *wh*-element in its specifier.

At this point, the natural question arises as to what is the exact nature of complementizer *of*. Although the claim that *of* is licensed by an (overt) *wh*-element in its specifier, i.e. an overt *wh*-phrase or an empty yes/no operator (cf. Zwart 2000 a.o.), accounts for its distribution in questions and RCs, it has nothing to say about the fact that *of* also functions as the Dutch disjunction marker. The different functions of *of* are illustrated in (227).

(227) a. Ze vroeg **of** hij het gedaan heeft.
she asked whether he it done has
'She asked whether he has done it.' *yes/no Q*

b. % Ze vroeg wie **of** het gedaan heeft.
she asked who whether it done has
'She asked who has done it.' *embedded wh-Q*

c. % Dat is de man wie **of** het gedaan heeft.
that is the man who whether it done has
'That is the man who has done it.' *restrictive RC*

d. Jan **of** Piet
Jan or Piet *disjunction*

Jayaseelan (2008) claims that the question operator is to be identified with the disjunction operator (cf. Bayer 2004), and he suggests that this identification holds universally.[36] In Dutch, *of* is taken to be the lexical realization of this disjunction/question operator. More specifically, *of* may function as a disjunction marker (cf. (227d)) and as a question particle (cf. (227a)). Jayaseelan (2008) argues that only if the disjunction/question operator *of* is merged in the CP domain it can output question semantics. In light of the split CP domain in Dutch as outlined above, this claim can be made a bit more specific by stating that *of* can lexicalize the question operator and output question semantics only in the higher CP layer, namely CP_1 (ForceP), of a clause with interrogative force.

Recall that *of* may also appear in RCs that are introduced by a relative pronoun with *wh*-morphology (cf. (227c)), in which case *of* occupies a position that is different from the position it occupies in questions: C^0_2 in RCs and C^0_1 in questions. In RCs, *of* in C^0_2 cannot output question semantics, i.e. it does not contribute anything to the semantics of the clause. Put differently, in RCs *of* appears to be merely a meaningless dummy element that can optionally be added to the clause. That this line of reasoning is on the right track, is evidenced by the fact that DFC with *of* in RCs is generally optional and does not have a clear geographic distribution (colloquial Dutch). The same holds for *of* in *wh*-Qs (cf. SAND1 data: Barbiers et al. 2005, Boef to appear), which seems to indicate more generally that *of* in DFC configurations is a meaningless dummy element.[37]

In sum, the element *of* in Dutch can apparently be two things: (i) the lexicalization of the disjunction/question operator (*of* as disjunction marker and *of* in yes/no questions), and (ii) a dummy complementizer that appears as the head of a projection with an overt *wh*-element in its specifier (*of* in *wh*-Qs and RCs introduced by a *wh*-pronoun).[38] Most likely, the use of *of* in DFC configurations developed from the yes/no question operator *of*, cf. Cheng and Rooryck (2000) (*of* in yes/no questions => *of* in *wh*-questions => *of* in RCs introduced by a *wh*-pronoun). I leave this for further research.

[36] Several scholars have argued that question semantics involves disjunction (see e.g. Higginbotham and May 1981, Higginbotham 1993).

[37] DFC with *dat* (in combination with a *wh*-pronoun) on the other hand is obligatory in most locations in which it is attested (southern Dutch), cf. SAND1 data (Barbiers et al. 2005). See Boef (to appear) for some discussion on the difference between *obligatory* and *optional* DFC.

[38] The simpler alternative approach to the nature of *of* – according to which *of* does not express any semantics and is only specified as [+wh], as a result of which it can only be licensed in a position the specifier of which is [+wh] (the null operator in yes/no questions then needs to be specified as [+wh]) – does not have anything to say about *of* as a disjunction marker. The observation that there are several unrelated languages that show homophony between the disjunction marker and the question particle (cf. Jayaseelan 2008) then becomes a mere coincidence.

3.4.4.2 Free relative clauses

If we take seriously the parallel between restrictive RCs and embedded *wh*-Qs introduced by *welke* 'which', the proposed structure for restrictive RCs in (226) predicts that in RCs that do not have a RC head, the higher SpecCP position should be available for complex *wh*-elements introduced by *welke*. Furthermore, DFC with *of dat* should be possible in such RCs, as *of* can be licensed in C_1^0 by the *wh*-phrase in its specifier (without making a contribution to the semantics). Abstracting away for the moment from the precise analysis of the syntax and semantics of free relative clauses (FRCs), on a descriptive level this seems to be the pattern that we find in Dutch FRCs, as illustrated in (228).

(228) a. **Welke onverlaat** zoiets doet krijgt straf.
which miscreant such a thing does gets punishment
'Whichever miscreant does such a thing will be punished.'
[= (179a), de Vries 2004:193]

b. % **Wie of dat** het weet mag het zeggen.
who whether that it knows may it say
'Who knows it may say it.' [MPQ1-A data]

The sentence in (228a) shows that Dutch FRCs may feature complex *wh*-phrases introduced by *welke*, in contrast to restrictive RCs.[39] The prediction that complex *wh*-phrases introduced by *welke* and RC heads in restrictive RCs are in complementary distribution, thus seems to be borne out by the data. The sentence in (228b) shows that in FRCs the string *wh of dat* is attested (albeit rather marginally; the MPQ1-A data show that between 13% and 15% of the informants accept the string *wie of dat* in FRCs). The observation that *of dat* is *not* attested in restrictive RCs can be accounted for in two ways: (i) the RC head cannot license the presence of *of* in C_1^0 (but see section 3.4.4.3) – in contrast to the *wh*-pronoun in FRCs as in (228b) – and (ii) there is not enough space to the right of the relative pronoun (in $SpecCP_2$) to fit both complementizers (i.e. there is only C_2^0).

[39] It should be noted that there are some restrictions on the presence of *welke* 'which' phrases in FRCs. Whereas 'regular' FRCs have a definite or universal interpretation, so-called *head internal FRCs*, like the one in (228a), can only get a universal interpretation, as observed by (de Vries 2004), and illustrated below.

(i) a. Welke bakker zo'n grote winkel heeft, zal vast wel witbrood verkopen.
which bakery such a big store has will probably white bread sell
'Whichever bakery has such a big store probably sells white bread.'
[universal]

b. * Welke bakker hier op de hoek zit, zal vast wel witbrood verkopen
INTENDED: 'Which bakery is here at the corner, probably sells white bread'
[definite, de Vries 2004:196]

Furthermore, *epithets* (like *welke idioot* 'which idiot' or *welke onverlaat* 'which miscreant') seem to occur in this environment most naturally, cf. appendix A to chapter 2.

Similarly, non-restrictive/appositive RCs allow – next to the antecedent – an additional internal head (introduced by *welke*) that introduces the RC itself, as illustrated in (229), cf. (224) *supra*.

(229) "In de ban van de ring", welk boek van Tolkien zeer populair is,
in the spell of the ring which book by Tolkien very popular is
is verfilmd.
has been filmed
'"The lord of the rings", which book by Tolkien is very popular, has been filmed.' [= (224b), de Vries 2004:193]

It is well known that there are many differences between restrictive and non-restrictive RCs (e.g. Fabb 1990). Such differences are often implemented as the result of different syntactic analyses. De Vries (2002, 2004) and Cheng and Downing (2006) amongst others claim that the RC head is base-generated *outside* the CP in non-restrictive RCs, whereas the RC head is base-generated *inside* the CP in restrictive RCs, just as in my analysis of restrictive RCs in (226) (in which the RC head is base-generated in SpecCP_1). Under such an analysis the difference between restrictives and non-restrictives regarding the potential presence of complex *wh*-phrases introduced by *welke* is easily accounted for. A *welke* phrase is only grammatical in non-restrictive RCs, as only in non-restrictive RCs is the SpecCP_1 position available: the RC head is base-generated outside the RC – as opposed to restrictive RCs in which the RC head is base-generated inside the RC (in SpecCP_1). If true, we furthermore predict that DFC with *of dat* should occur in non-restrictive RCs. This prediction needs to be tested.

3.4.4.3 Amsterdam Dutch

As reported by Hoekstra (1994), Amsterdam Dutch allows a *wh*- or *d*-pronoun in RCs to be *preceded* by complementizer *of*, as illustrated in (230).

(230) a. de auto **of** **waar** ik in reed
the car whether where I in drove
'the car in which I drove'

b. de vrouw **of** **die** ik gezien heb
the woman whether RP I seen have
'the woman who I have seen'
[Amsterdam Dutch, Hoekstra 1994:316]

The Amsterdam Dutch data can be accommodated by the structure in (226) by arguing that the element *of* in (230) is a relative complementizer or clause typer (C_{REL} in C^0_1). That this analysis of *of* in Amsterdam Dutch might be on the right track is provided by the generalization that normally *of* in Dutch RCs can only be licensed by an (overt) *wh*-element in its specifier (cf. section 3.4.4.1). That *of* in (230) does not adhere to this generalization – *of* precedes

the relative pronoun and *of* may occur with a *d*-pronoun – suggests that it is a different element. The Amsterdam Dutch data thus seem to provide evidence for a head position (C^0_1) in between the RC head and the RC itself.

A proviso is in order here: the status of the data in (230) is unclear. The Amsterdam Dutch data have been tested in MPQ1-A, but they were found to occur only very marginally. To the extent that they exist at all, only test sentences with the *wh*-pronouns *wat* 'what' and *waar* 'where' (in locative RCs) are attested, as illustrated in (231). The geographic distribution of this phenomenon (if any) is not restricted to (the area of) Amsterdam; the term *Amsterdam Dutch* is thus clearly a misnomer.

(231) a. ?* Het boek **of** **wat** ik gelezen heb is mooi geschreven.
the book whether wat I read have is beautifully written
'The book that I read is beautifully written.' [29/452=6%]

b. ?* Het huis **of** **waar** ik gewoond heb is verkocht.
the house whether where I lived have is sold
'The house where I lived is sold.' [15/452=3%, MPQ1-A data]

Assuming that the Amsterdam Dutch data indeed exist, the structure in (226) in fact predicts there to be dialects in which all positions in the left periphery of RCs are spelled out: RC head ($SpecCP_1$) – *of* (C^0_1) – relative pronoun ($SpecCP_2$) – *dat* (C^0_2). The Amsterdam Dutch data in (231) have been tested with an additional declarative complementizer. As illustrated in (232), such patterns were not attested in MPQ1-A (less than 2% of the informants accepted such sentences). The lack of attestations of this construction casts further doubt on the reality of the Amsterdam Dutch data in (230), i.e. RCs in which the relative pronoun is *preceded* by complementizer *of*.

(232) a. * Het boek **of** **wat** **dat** ik gelezen heb is mooi geschreven.
the book whether wat that I read have is beautifully written
'The book that I read is beautifully written.'

b. * Het huis **of** **waar** **dat** ik gewoond heb is verkocht.
the house whether where that I lived have is sold
'The house where I lived is sold.' [MPQ1-A data]

3.4.4.4 Double complementizers

The claim that the C^0_1 position is targeted by elements that are specific to RCs receives some additional support from the observation that there are languages that make use of relative complementizers. There is evidence that the relative complementizer in such languages does in fact target the higher C^0_1 layer. The relative complementizer – which I argued targets the higher C^0_1 – can optionally be followed by the finite declarative complementizer – which I argued targets the lower C^0_2. This is illustrated for Slovene and Icelandic in (233).

(233) a. človek, [$_{CP1}$ ki [$_{CP2}$ (da) pride]]
the man C_{REL} that is coming
'the man that is coming' [Slovene, Hladnik 2010:14]

b. manninn [$_{CP1}$ sem [$_{CP2}$ (að) kom hingað]]
the man C_{REL} that came here
'the man that came here' [Icelandic, Boef and Franco 2010]

The examples in (233) thus seem to indicate that the relative complementizer targets a position in the left periphery that is higher than the position targeted by the declarative complementizer, in line with the proposal in (226).[40]

3.5 Rethinking the HEA of relative clauses

In the previous sections, I have argued for a specific implementation of a head *external* analysis of restrictive RCs in Dutch, to accommodate (among other things) the colloquial Dutch doubling data as discussed in chapter 2. In this section, I will argue in more detail in favor of a HEA of RCs more generally. That is to say, I will make a case for the HEA and argue against HIAs (*raising* or *matching*) of RCs. I will show that although HIAs have featured most prominently in syntactic theorizing in recent years, (my implementation of) the HEA (as introduced and discussed in the previous section, cf. (226)) fares better in many respects: case mismatches (section 3.5.1), selection (section 3.5.2), and locality constraint violations (section 3.5.3). I will furthermore briefly mention coordination and extraposition in the domain of RCs in section 3.5.4. A detailed evaluation of the most prominent argument in favor of HIAs of RCs – namely connectivity effects – will be given in section 3.6.[41]

[40] One might expect that languages can only have *complementizers* that are specific to RCs, not *relative pronouns*, as a relative pronoun occurs in the lower CP domain, which is not related to clause typing (cf. section 3.4.3). However, according to the typology in de Vries (2002), this prediction is not entirely borne out. See chapter 4 for discussion on the nature of relative pronouns. For now it suffices to say that it is conceivable that languages may differ in their exact make up of the left periphery (for a parametric account of the size and the properties of the left periphery see e.g. Bhatt and Yoon 1992).

[41] Internally headed RCs (or RCs that contain multiple overt heads, cf. Cinque 2010 for Italian and Latin), as illustrated in (i), have been taken as strong evidence in favor of HIAs of RCs: they seem to display the structure before *raising* (cf. de Vries 2002:77). My aim is not to argue against HIAs of RCs altogether as it might very well be the case that it is necessary to distinguish between different types of RCs, both within and across languages. Rather, my more modest aim is to make a case for the (particular implementation of the) HEA of RCs (as outlined in (226)). Accordingly, I will not be concerned with internally headed RCs in this thesis.

(i) [nuna bestya-ta ranti-shqa-n] alli bestya-m ka-rqo-n
man horse-ACC buy-PERF-3 good horse-EVIDENTIAL be-PAST-3
'The horse that the man bought was a good horse.'
[Ancash Quechua, Cole 1987:277]

3.5.1 Case mismatches

Because both the *raising* analysis and the *matching* analysis of RCs take the RC head to be merged as the complement of the relative pronoun (D^0) inside the RC, the case of the RC head is predicted to be compatible with its grammatical function inside the RC, and with the case of the relative pronoun. However, in languages with overt case morphology, case mismatches between the relative pronoun and the RC head are attested, as exemplified for Polish in (234) and for German in (235).[42]

(234) Widziałem tego pana, [który zbił ci szybę].
saw.1SG the.ACC man.ACC who.NOM broke your.SG glass.ACC
'I saw the man who broke your glass.' [Polish, Borsley 1997:638]

(235) Wir brauchen den Politiker, der unsere Interessen vertritt.
we need the.ACC politician who.NOM our interests represents
'We need the politician who represents our interests.'
[German, Schmitt 2000:332]

Although several solutions to the case mismatch problem have been proposed for the HRA (cf. Kayne 1994, Bianchi 2000, de Vries 2002), they are all (to a greater or lesser extent) stipulative and not particularly explanatory (Salzmann 2006:17).[43] For the MA of RCs, the case mismatch data are less problematic

[42] I am abstracting away here from so-called *case attraction* phenomena: situations in which the RC head bears the same case as the relative pronoun. This is either the RC internal case as in (i): (regular) *case attraction* (i.e. the RC head bears the case of the relative pronoun), or the RC external case as in (ii): *inverse case attraction* (i.e. the relative pronoun bears the case of the RC head). All examples in (i)-(ii) are from Bianchi (1999:93-95). I will not be concerned with case attraction as it only seems to be at play in free relatives and correlatives (Bianchi 1999, Bhatt 2005), i.e. case attraction does not seem to occur in externally headed *restrictive* RCs – the type of RC I am concerned with.

(i) *case attraction*

a. notante iudice quo nosti
judging.ABL the judge.ABL whom.ABL (you) know
'judging the judge whom you know' (ACC => ABL) [Latin, Hor., *Sat* 1,6,15]

b. sie gedâht' ouch maniger leide, der ir dâ héimé geschach.
she thought also some sufferings.GEN which.GEN her at home happened
'She thought of some pain that she suffered at home.' (NOM => GEN)
[Old High German]

(ii) *inverse case attraction*

a. Urbem quam statuo vestra est.
city.ACC which.ACC found yours is
'The city that I found is yours.' (NOM => ACC) [Latin, Verg. *Aen.* I, 573]

b. Den schilt den er vür bôt der wart schiere zeslagen
the.ACC shield.ACC which.ACC he held that.NOM was quickly shattered
'The shield that he held was quickly shattered.' (NOM => ACC)
[Old High German]

[43] Bianchi (2000:129-130) assumes that a N^0 receives the same case as the D^0 by which it is *governed*. Since the RC head ends up being governed by the external D^0 and not by the

because the external head is not in a chain relation with the internal head, so they can receive case separately as long as total identity of features is not required for *deletion under identity* (cf. Citko 2001). Only from a traditional HEA of RCs do case mismatches like (234) and (235) follow straightforwardly, as there is no RC internal representation of the RC head that receives case RC internally. For the proposed implementation of the HEA of RCs in (226), something additional needs to be said about the case mismatch data. More specifically, whereas the HEA in (226) correctly predicts the RC head not to bear RC internal case (unlike HRAs without additional assumptions, cf. footnote 43), it is unclear how exactly case on the RC head is licensed in this structure and how case on the RC head (in SpecCP_1) can be different from case on the relative pronoun (in SpecCP_2). While it is beyond the scope of this thesis to engage in an in-depth analysis of case in RCs – in part because there is no visible (morphological) case on determiners, nouns and relative pronouns in Dutch – let me point out that, by whatever mechanism, case on the relative pronoun must be licensed RC internally by the verb, whereas case on the RC head must be licensed by the external D^0.[44] I leave the exact analysis of these case relations for further research (see Schmitt 2000 for case in a RC structure like (226)).

A related problem for the *raising* analysis comes from *adjectival inflection*, as pointed out by Heck (2005:3) and Salzmann (2006:123-124) among others. The form of adjectives in attributive position of neuter nouns (e.g. *huis* 'house') in Dutch depends on the form of the determiner. When the definite determiner *het* 'the' is used, the default *-e* appears, and when the indefinite determiner *een* 'a(n)' is used, the adjective appears without an (overt) agreement affix (cf.

relative D^0, the RC head shows the same morphological case as the external D^0, whereas the relative pronoun shows RC internal case (case is thus assumed to be determined after syntax, in the morphological component). Borsley (2001) notes that this account of case in RCs runs into several problems, among which that it predicts that in D+CP structures without NP but with *wh*-movement of the *wh*-pronoun, the external determiner and *wh*-pronoun bear the same case as the pronoun is governed by the external determiner. This prediction is not borne out, as illustrated in (i) for Polish (= (199)).

(i) To, [$_{CP}$ kogo Maria widziała] jest tajemnicą.
that.NOM who.ACC Maria saw is secret
'Whom Mary saw is a secret.' [Borsley 1997:631]

Alternatively, de Vries (2002) argues that the formal features of the RC head incorporate into the external determiner in order to check its case feature. The external D^0 and the RC head thus eventually end up bearing the same case, and their case may differ from the case on the relative pronoun. As correctly pointed out by Salzmann (2006:15-17), both solutions to the case mismatch problem remain descriptive as they are only needed to save the HRA of RCs, i.e. the assumptions they make are not independently motivated but are only at play in the empirical domain of RCs.

[44]The external D^0 thus needs to be able to interact with the specifier of its complement. According to current minimalist assumptions the edge of a phase (specifier) can be reached from outside that phase (*Phase Theory* and *Phase Impenetrability (PIC)*, Chomsky 2000 *et passim*). Since the RC head is on the edge of a phase (SpecCP_1), the RC head is thus available for further computation after the CP is built.

section 2.5.2.1). This is illustrated in (236), cf. example (64).

(236) a. het mooi-e huis
the beautiful house

b. een mooi huis
a beautiful house

As can be seen in (237), in Dutch RCs, the form of the adjective depends on the external D^0, not on RC internal material. This follows from a HEA of RCs, because the adjective is base-generated outside the RC. A *raising* analysis on the other hand, incorrectly predicts that the adjective always has the default *-e* ending because the neuter RC head *huis* 'house' combines with the definite demonstrative relative pronoun *dat* 'that' inside the RC: *dat mooi-e huis* 'that beautiful house'.[45]

(237) a. het mooi-e/*mooi huis dat zij gekocht hebben
the beautiful house that they bought have
'the beautiful house that they have bought'

b. een mooi/*mooi-e huis dat zij gekocht hebben
a beautiful house that they bought have
'a beautiful house that they have bought'

3.5.2 Selection

In contrast to what we predict on the basis of an analysis that takes relative pronouns to be determiners – as in a HIA of RCs – relative pronouns do not have the same selectional properties as their determiner or interrogative counterparts. This is exemplified in (238) for the Dutch pronoun *die* (= (75) from section 2.5.2.4).

(238) a. <dat/*die> meisje heb ik gezien
that.N/that.C girl.N have I seen

b. het meisje <dat/%die> ik gezien heb
the girl.N that.N/that.C I seen have

This argument from selection becomes even stronger when we consider the fact that some pronouns that can act as relative pronouns cannot act as interrogative determiners, as exemplified for the English pronoun *who* in (239).

(239) a. the man [who I have seen]

b. * I have seen [who man]

c. * [who man] have you seen?

[45] Alternatively, one could assume that the prenominal adjective in RCs like (237) is not part of the constituent that contains the RC head. Following the line of reasoning in Kayne (1994) and Bianchi (2000), the prenominal adjective could be in SpecCP, whereas the RC head plus RC is in SpecIP. Although that solves the problem that comes from adjectival inflection, such an analysis runs into other problems as discussed in Borsley (2001).

Similarly, under a HIAs of RCs, adverbial RCs like (240) would have to base-generate [waarom reden] ('[why reason]'). This seems implausible (cf. Aoun and Li 2003:121; Salzmann 2006:14,79-80), because *waarom* normally cannot act as an interrogative determiner (**waarom reden* 'why reason').

(240) de reden waarom hij niet kwam
the reason why he not came
'the reason why he did not come'

Obviously, this problem of selection is not insurmountable as one can always assume that the relative pronoun simply is a different lexical item than its interrogative or determiner counterpart, i.e. there are (at least) two lexical entries for Dutch *die* and English *who*. Such a conclusion is theoretically unattractive because it would result in the postulation of several construction-specific lexical entries (cf. Wiltschko 1998). Moreover, it cannot account for the fact that the meaning and the categorial status of a pronoun is (partially) determined by the syntactic configuration it occurs in (cf. Postma 1994, Cardinaletti and Starke 1999, Koopman 1999 a.o.).[46,47] See chapter 4 for detailed discussion on the nature of relative pronouns, and on pronouns that can have more functions (*multipurpose pronouns*) more generally.

A further argument against HIAs of RCs concerns the external determiner in English non-*wh*-relatives like (241).

(241) He bought *every* book that he liked.

As argued by Hackl and Koster-Moeller (2008) and Koster-Moeller (2012), the external determiner *every* also originates inside the RC, as indicated in (242).[48]

(242) He bought every book that he liked <*every* book>.

Such a proposal runs into interpretation difficulties related to selection as *he liked every book* certainly does not mean *every book he liked*. To solve this problem, we could assume that the external determiner *every* is merged outside the RC – as is the common assumption – and that only the noun *book* moves from a RC internal position to a position in which it becomes the RC head (it

[46]The selection problem pointed out for HIAs of RCs might not be so problematic for *raising* analyses: although the relative pronoun is merged in the same position as its determiner or interrogative counterpart, it eventually ends up in a different position. This might account for the different properties of relative pronouns on the one hand and its determiner or interrogative counterparts on the other hand.

[47]As was already pointed out in section 2.5.2.4, under a HEA of RCs, the pattern in (238) is in line with the universal pattern of agreement captured by the *Agreement Hierarchy* proposed by Corbett (1979, 1991, 2006).

[48]This proposal can only hold for non-*wh*-relatives and not for relatives involving (*wh*-) pronouns. According to HIAs of RCs the relative (*wh*)-pronoun acts as a determiner selecting for the RC head, and there cannot be two determiners selecting for the RC head. This observation is problematic in itself, as there does not appear to be any interpretational difference between e.g. *He bought every book* ***that*** *he liked* and *He bought every book* ***which*** *he liked*.

is irrelevant for present purposes what that position is), as illustrated in (243). However, this results in a number mismatch, as the RC head can only be properly merged with *liked* when it is plural – as indicated in (244), the source position is incompatible with a bare noun – whereas it needs to be singular when it is combined with the external determiner *every* (**every books*), cf. Borsley (1997, 2001).

(243) He bought every book that he liked <book>.

(244) a. He liked books.
b. *He liked book.

Obviously, selectional problems like the ones outlined above do not arise when a HEA of RCs is assumed, because (i) the relative pronoun does not select the RC head as its complement, and (ii) the RC head (together with the relative pronoun) is not merged as an argument of the verb inside the RC.

In contrast, sometimes information about the RC head needs to be present RC internally, which has been used as an argument in favor of HIAs of RC. As noted by Bhatt (2002), Larson (1985) observes that bare NP adverbs of manner, e.g. *way* and *fashion*, behave quite differently (which cannot be attributed to their semantics as they are close synonyms). More specifically, only NPs headed by *way* are capable of appearing as bare NP adverbs, as illustrated in (245).

(245) a. you pronounced my name that way
b. * you pronounced my name this fashion [Larson 1985:598]

Interestingly, the same contrast between *way* and *fashion* (or *manner*) holds in adverbial RCs (of manner), i.e. such RCs are only grammatical with *way* as the RC head.

(246) a. the way (that) you talk
b. * the fashion/manner (that) you talk [Larson 1985:616]

Bhatt (2002) claims that under a HIA of RCs the ungrammaticality of (246b) directly follows from the ungrammaticality of (245b): in (246b) the RC head *fashion/manner* is assumed to be merged inside the RC as the complement to the verb *talk*, but we know independently that that is not possible from (245b). A HEA on the other hand cannot directly account for the fact that information about the RC head is present inside the RC. Some additional mechanism of *feature transmission* is needed to make the relevant information about the RC head visible inside the RC (as proposed by Larson 1985). However, under a HEA of RCs, the RC head and the relative pronoun or operator are related through predicate modification (cf. section 3.4.3), as a consequence of which they will automatically share some features (cf. Salzmann 2006). Although the details of this kind of feature transmission in RCs need to be worked out, the argument from *subcategorization* against the HEA is thus not very well founded.

3.5.3 Locality constraint violations

Head *internal* analyses of RCs, and most prominently *raising* analyses of RCs, violate several well-established (locality) constraints on movement. The main sources for such arguments against HIAs of RCs that I will mention in this section are Borsley (1997, 2001), Salzmann (2006) and Webelhuth (2011).

Most versions of a *raising* analysis – though not the analysis by Kayne (1994), cf. *infra* – in one way or another violate the Condition on Extraction Domain (henceforth CED, Huang 1982) or the Freezing Principle (Wexler and Culicover 1980), which (simply put) state that a phrase that has undergone movement becomes an island for extraction. Because under a HRA of RCs the RC head is base-generated as the complement of the relative D^0 (relative pronoun or empty operator; see section 3.5.2), movement needs to take place in order to get the RC head in a position linearly preceding the relative D^0. As noted in section 3.2.2, one way of doing this is by moving the RC head out of the relative DP to (an extended projection of) CP (e.g. Zwart 2000, and Bianchi 1999, 2000 for *wh*-relatives) or to a position outside CP (e.g. Vergnaud 1974). Moving the RC head out of the DP_{REL} – after DP_{REL} has moved to the left periphery, as illustrated in (247) – violates the CED.[49]

(247) the man$_2$ $[_{DPREL}$ who $\__2]_1$ they have called $\__1$

De Vries (2002) – in part following Kayne (1994) – proposes to overcome this CED violation by assuming that the RC head moves from the complement position to the specifier position of DP_{REL} (before the DP_{REL} itself moves up to the left periphery, to avoid *countercyclicity*). According to some theories, this movement is too local (*Anti-Locality*, e.g. Grohmann 2003, Abels 2003). In addition, de Vries (2002) still needs to assume that the formal features of the RC head move out of DP_{REL} to the external determiner (to check case, cf. *supra*). This movement operation (albeit formal feature movement), as illustrated in (248), violates the CED (cf. Salzmann 2006). Notice that CED violations are non-existent given a HEA or MA of RCs, as there is no extraction of the RC head (out of DP_{REL}) at any point in the derivation.

(248) the+FF$_2$ $[_{DPREL}$ man$_2$ who $\__2]_1$ they have called $\__1$

A similar argument against most HRAs of RCs comes from extraction out of adjunct PPs in German. As mentioned by Webelhuth (2011), the NP-complement of an adjunct PP in German cannot be extracted, as illustrated in (249).

[49] Besides subextraction of the RC head being a violation of the CED, there is little consensus about the trigger for this particular movement operation (which often relies on construction-specific mechanisms, cf. Salzmann 2006:14ff.). For Bianchi (1999, 2000) and de Vries (2002) a.o. this step is triggered by the need to enter into a checking or agreement relation with the external determiner. For Zwart (2000) the trigger is semantic: subextraction of the RC head is triggered by the need to get the right constituency for *set intersection* of the RC head and the RC itself (cf. footnotes 33 and 34).

(249) * [Welchem Tag]$_1$ hatte Petra [$_{PP}$ an $\__1$] Urlaub?
which day had Petra on vacation
'Which day did Petra take off?' [Webelhuth 2011:24]

The example in (250) gives the RC counterpart of (249), which is perfectly grammatical. The contrast in grammaticality between (249) and (250) is unexpected under a HRA: in both cases the complement of an adjunct PP is extracted. Assuming a HRA of RCs thus leads to losing a generalization, namely that adjunct PPs are islands for extraction (cf. Heck 2005).[50,51,52]

(250) der Tag$_2$, [$_{PP}$ an dem $\__2$]$_1$ Petra $\__1$ Urlaub hatte
the day on which Petra vacation had
'the day that Petra took off' [Webelhuth 2011:24]

An argument against both *raising* and *matching* comes from the syntax of adpositions in German (Webelhuth 2011:26-27). "Adpositions of the *mit*-class are obligatorily prepositional if they combine with an inanimate phrasal complement and obligatorily postpositional if they combine with an inanimate pronominal complement," as illustrated in (251).

(251) a. Wir hatten [$_{PP}$ <mit> dem Anruf <*mit>] gerechnet.
we had with the call with expected
'We had expected the phone call.'

b. Wit hatten [$_{PP}$ <*mit> da <mit>] gerechnet.
we had with it with expected
'We had expected it.' [Webelhuth 2011:26]

Whereas interrogative clauses respect this generalization, as illustrated in (252), under a HIA of RCs, RCs violate this generalization, as can be seen in (253):

[50]The PP adjunct extraction cases are a subcase of the CED.

[51]A similar argument holds for Dutch. Dutch is a *partial* preposition stranding language, in the sense that it generally only allows prepositions to be stranded by R-pronouns or empty operators (van Riemsdijk 1978). Under most HRAs of RCs, in case of PP relativization the RC head moves out of the PP, as illustrated in (i). This movement violates the CED, as well as the condition on preposition stranding – although (i) strictly speaking does not show preposition stranding as the relative pronoun is stranded as well.

(i) de man$_2$ [met wie $\__2$]$_1$ ik $\__1$ gesproken heb
the man with who I spoken have
'the man I spoke with' [Salzmann 2006:19]

[52]Let me point out that the raising analysis as proposed by Kayne (1994) does not encounter this problem. The string *der Tag an dem* in (250) is most likely analyzed as in (i), cf. Kayne (1994:89). The RC head *Tag* moves to SpecPP (instead of out of the PP), and thus does not violate the CED. Notice that the movement of the RC head in (i) is *countercyclic*. However, this is easily overcome by assuming that the RC head moves to SpecPP first, after which the whole PP moves to SpecCP (cf. de Vries 2002).

(i) der [$_{CP}$ [$_{PP}$ Tag$_2$ [an dem $\__2$]]$_1$ ... $\__1$ ...

before movement, the adposition combines with an inanimate phrasal complement so we expect it to be prepositional (analogously to (251a)), but instead it is postpositional. If we were to assume a HIA of RCs (*raising* or *matching*), we would thus have to assume that a well-established constraint does not hold for RCs.[53]

(252) [$_{PP}$ <*mit> wo <mit>] hattet ihr nicht gerechnet?
with what with had you not expected
'What did you not expect?' [Webelhuth 2011:26]

(253) Etwas Schreckliches$_1$, [$_{PP}$ <*mit> [wo _$_1$] <mit>] wir nicht
something terrible with what with we not
gerechnet hatten
expected had
'something terrible that we had not expected to happen'
[Webelhuth 2011:27]

The last problem I want to mention here comes from possessive RCs. As observed by Borsley (2001) and Bhatt (2002) amongst others, moving the RC head out of the possessive phrase in SpecCP is an unusual extraction. More specifically, under most *raising* analyses of RCs, the relevant part of (254a) most likely has the underlying structure as given (somewhat simplified) in (254b).[54]

[53] Although the same argument can be made for Dutch, Dutch behaves slightly differently from German. Whereas the same generalization seems to hold – adpositions are obligatorily prepositional if they combine with an inanimate phrasal complement and obligatorily postpositional if they combine with an inanimate *d*-pronoun, as in (i) – adpositions combined with an inanimate *wh*-pronoun (*wat* 'what') can be both prepositional and postpositional in *wh*-clauses (iia), but must be postpositional in RCs (iib). However, when an adposition is combined with an animate *wh*-pronoun (*wie* 'who'), in *wh*-Qs it needs to be prepositional to get the animate interpretation of the *wh*-pronoun (iiia), whereas in RCs it can be prepositional as well as postpositional (iiib) – the latter construction is mostly found in informal Dutch (Haeseryn et al. 1997:344).

(i) a. we hebben [met het speelgoed (*met/mee)] gespeeld
we have with the toys with played

b. we hebben [*met dat] / [daarmee] gespeeld
we have with that there with played

(ii) a. [met wat] / [waarmee] hebben we gespeeld?
with what where with have we played

b. het speelgoed [*met wat] / [waarmee] we hebben gespeeld
the toys with what where with we have played

(iii) a. [met wie] / [#waarmee] hebben we gespeeld?
with who where with have we played

b. de man [met wie] / [waarmee] we hebben gespeeld
the man with who where with we have played

[54] Again, this argument does not hold for the raising analysis as proposed by Kayne (1994), as in such an analysis the RC head moves to the specifier position of the *wh*-phrase, and thus never moves out of the possessive noun phrase.

Moving the RC head from the possessive noun phrase, as illustrated in (254c), violates the CED/Freezing Principle, and more importantly, it violates the ban on extraction from a possessive specifier, which otherwise holds for English (cf. Borsley 2001: **Which country did you meet the president of's wife?*).[55] The same holds for Dutch, as illustrated in (255) (cf. Webelhuth 2011:21 for German) – notice that we are basically dealing with a Left Branch Condition violation in (255a).

(254) a. the man whose father did it
b. [[whose man] father]
c. man$_2$ [[whose _$_2$] father]$_1$ ... _$_1$

(255) a. * Wiens$_1$ heeft [_$_1$ vader] het gedaan?
whose has father it done
INTENDED: 'Whose father has done it?'
b. de man$_1$ [wiens _$_1$ vader] het heeft gedaan
the man whose father it has done
'the man whose father has done it'

Example (256) illustrates the related problem of 'unbounded possessor extraction' (cf. Bhatt 2002:81-83): extraction of unboundedly deeply embedded possessors. Put differently, extraction of the RC head from a position deeply embedded inside the possessive noun phrase is unorthodox.

(256) the student$_2$ [[[whose _$_2$] brother's] band]$_1$ Jonah likes _$_1$
[adapted from Bhatt 2002:81]

As noted by Bhatt (2002:81-83) (cf. Salzmann 2006:18) – following Åfarli (1994) a.o. – the raising analysis is most likely not available in cases that involve complex pied-piping.[56] Both the HEA and the MA of RCs can account for possessive RCs without any additional stipulations as the RC head is never extracted from the possessive DP.

[55]The violation of the ban on extraction from a possessive specifier is a subcase of the CED (cf. Salzmann 2006).

[56]De Vries (2002, 2006) proposes an account of possessive RCs under a promotion analysis of RCs (raising + D-complement). Although this analysis successfully combines the syntax of attributive possession with the syntax of relativization (in terms of raising), it still faces a violation of the CED/Freezing Principle, as well as a violation of the ban on extraction from a possessive phrase (in terms of formal feature movement, cf. *supra*). As illustrated in (i)-(ii), the final step in the derivation of a phrase like *de man wiens vader ik ken* 'the man whose father I know' involves movement of the formal features of the RC head *man* to the external determiner for ϕ-feature checking.

(i) [$_{DP}$ [$_{PP}$ man$_2$ D$_{rel,FF1}$+P$_{gen}$ [$_{DPrel}$ [$_{Drel}$ wiens]$_1$ _$_2$]]$_3$ D$_{poss}$ [$_{NP}$ vader _$_3$]]
(ii) de+FF$_2$ [$_{CP}$ [$_{DP}$ man$_2$ wiens vader]$_1$ ... ik ken _$_1$]

3.5.4 Coordination and extraposition

Extraposition and coordination facts show that the RC is a constituent to the exclusion of the RC head (and the external determiner), as illustrated in (257) for extraposition. This is compatible with the (traditional) HEA and the MA of RCs, but it is problematic for HRAs that do not assume that the RC head moves out of DP_{REL} (Kayne 1994, de Vries 2002), as under such analyses the RC head forms a constituent with the RC to the exclusion of the external determiner, cf. Borsley (1997).[57,58]

(257) Ik heb [$_{RChead}$ *de man*] gezien [$_{RC}$ *die zijn tas verloor*].
I have the man seen RP his bag lost
'I have seen the man who lost his bag.' [de Vries 2002:233]

In chapter 7 of his dissertation, de Vries (2002) gives a detailed empirical and theoretical evaluation of *all* existing theories of extraposition *in the broad sense* (i.e. independent of extraposition in RCs), like the *rightward movement* analysis (e.g. Reinhart 1980, Baltin 1984), the *stranding* analysis (e.g. Kayne 1994), and the *specifiying coordination* analysis (e.g. Rijkhoek 1998, Koster 2000a). He concludes that all empirical and theoretical issues are best accounted for by a *specifying coordination* account of extraposition, as in (258), and more specifically, a *specifying coordination plus ellipsis* account as in (259). The latter analysis is compatible with a HRA of RCs, cf. de Vries (2002:chapter7) for details.

(258) Ik heb [*de man* gezien [&: *die zijn tas verloor*]]]
specifying coordination

(259) Ik heb [*de man* gezien [&: ~~de man~~ *die zijn tas verloor* ~~gezien~~]]]
specifying coordination plus ellipsis

A theory of extraposition in RCs is thus completely dependent on a theory of extraposition, i.e. extraposition is a general phenomenon that not only applies to RCs. The fact that RCs can be extraposed is in itself thus not an argument in (dis)favor of any of the three analyses of RCs. Rather, which analysis of RCs should be favored is dependent on which analysis is best compatible with the

[57] Within the *Antisymmetry* framework (Kayne 1994), extraposition of postnominal RCs cannot be the result of rightward movement. Rather, extraposed RCs are the result of a leftward movement process of the RC head that leaves behind/strands the RC. There are many problems with such a theory (see a.o. Borsley 1997, de Vries 2002:chapter7).

[58] Under the HEA that I proposed in section 3.4, extraposition most likely involves (displacement of) CP_2. However, if true, languages that have a complementizer that is specific to RCs (e.g. Slovene, Icelandic, cf. section 3.4.4.4) and that allow extraposition, are predicted to have an extraposition structure as in (i), *quod non*. This suggests that given the RC structure in (226), extraposition most likely does not involve (movement of) CP_2, but rather (movement of) $\mathrm{C'}_1$.

(i) [[$_{CP1}$ RC head C_{REL}] ... [$_{CP2}$ C]]

analysis of extraposition one chooses. As the study of extraposition is a field of research on its own, it is certainly beyond the scope of this thesis to provide an account of extraposition *in the broad sense*. Furthermore, as each analysis of RCs can be made compatible with (a variant of) the analysis of extraposition in (259), at this point, I do not think that extraposition in RCs can be used to argue in favor of any analysis of RCs.[59]

In addition to the ability of being extraposed, RCs can have multiple heads, as illustrated in (260); the b-example is the Dutch equivalent of the English a-example. These coordination constructions are better known as *hydras*.

(260) a. the man and the woman [who were crazy]
b. de man en de vrouw [die gek waren]

Furthermore, as first observed by Ross and Perlmutter (1970), restrictive RCs can also have *split antecedents*. As exemplified in (261) and (262), the antecedent of such plural RCs in sentence-final position is a discontinuous noun phrase; the b-examples are the Dutch equivalent of the English a-examples.

(261) a. *A man* entered the room and *a woman* went out [who were quite similar]. [Ross and Perlmutter 1970:350]
b. *Een man* kwam binnen en *een vrouw* ging naar buiten [die er hetzelfde uitzagen].

(262) a. First *a man* arrived, then *a woman* arrived, and finally *a boy* arrived [who all looked like zombies]. [Hoeksema 1986:69]
b. Eerst arriveerde *een man*, toen arriveerde *een vrouw*, en uiteindelijk arriveerde *een jongen* [die er allemaal als zombies uitzagen].

These sentences suggest an analysis in terms of *NP adjunction*, as the external determiners take scope over the RC. However, the plural agreement on the verb suggests that the RC adjoins to the coordination of the RC heads: *DP adjunction*. Suñer (2001) proposes to overcome this paradox by assuming that each conjunct is modified separately by the RC, and that the first RC deletes (under identity with the second): *backwards deletion*, as illustrated in (263). Assuming furthermore that agreement is partially dependent on semantic interpretation, the pluralization of the verb (and of the relative pronoun in languages that show agreement on the relative pronoun) inside the RC can be accounted for.[60]

[59]The question of whether or not extraposed RCs show connectivity effects has always featured prominently in the discussion regarding extraposition in the domain of RCs. It has been claimed that there are no connectivity effects in extraposed RCs (e.g. Hulsey and Sauerland 2006), but the opposite has been suggested as well (de Vries 2002). Since I do not believe connectivity effects to be a reliable diagnostic for movement in the first place (see section 3.6), I do not think that the presence or absence of connectivity effects in extraposed RCs can tell us anything about which analysis of RCs – or, more generally, which analysis of extraposition – is to be favored.

[60]Similarly, RCs with collective predicates, as illustrated for the predicate *to meet* in (ia), need to get a semantic account under this approach to RCs with conjoined antecedents. More

(263) [$_{\&P}$ [the man ~~who was/were crazy~~] [$_{\&}$ and [the woman who was/were crazy]]]

Although it is unclear how exactly plural agreement comes about, notice that under this analysis of coordination phenomena it is irrelevant whether a HEA or a HIA is assumed. In sum, since coordination phenomena pose a big challenge for *any* analysis of the syntax of RCs – the most pressing question being where the plural agreement on the verb (and relative pronoun) inside the RC comes from – they cannot be used to argue in favor or against a particular analysis of RCs. As we saw above, the same holds for extraposition: it needs to be clear first what is the best account of extraposition (which is for the most part highly theory-dependent), before one can argue in favor or against a particular analysis or RCs on the basis of extraposition data.

3.5.5 Interim summary and outlook

If connectivity effects are treated in terms of syntactic reconstruction (i.e. interpreting a lower copy at LF, cf. Chomsky 1993, Fox 1999b a.o.), connectivity effects are a diagnostic for movement. For proponents of a syntactic reconstruction account to connectivity effects, the presence of connectivity effects in RCs has been the most important argument in favor of analyses that assume there is a representation of the external head inside the RC, i.e. head *internal* analyses of RCs. Given the observation that long-distance RC that show doubling manifest connectivity effects (*infra*), we are faced with the following paradox: doubling data, case mismatches, problems from selection, and locality constraint considerations favor a HEA of RCs, whereas connectivity effects challenge such an analysis. This is summarized in table 3.3 below.

Table 3.3: Summary properties three main analyses of RCs

phenomenon	HEA	HRA	MA
doubling in long-distance RCs	+	-	+/-
case mismatches/adjectival inflection	+	+/-	+
selectional differences between D_{REL} and $D_{(Q)}$	+	+/-	+/-
locality constraint violations	+	-	+/-
(anti-)connectivity effects	+/-	+/-	+

specifically, the requirement that a predicate like *to meet (in secret)* needs a plural subject is not fulfilled by the analysis of (ia) as given in (ib), as a result of which the grammaticality of (ia) needs to be accounted for *semantically* (see Suñer 2001:274-275 for details).

(i) a. The boy and the girl who met in secret were discovered by her parents.
b. [The boy ~~who met in secret~~] and [the girl who met in secret] were discovered by her parents. [Suñer 2001:274]

It thus seems wrong to only consider connectivity effects (between the RC head and the RC internal gap) in determining the right analysis of RCs. The remainder of this chapter focuses on connectivity effects in RCs and shows that they cannot always be treated by means of syntactic reconstruction, i.e. reconstruction without movement (or copies) seems to be independently needed anyway. Connectivity effects in RCs thus cannot be used as a proper diagnostic for movement of the RC head. Consequently, connectivity effects should not be used as (the only) diagnostic for distinguishing the HEA and HIAs of RCs.

3.6 Connectivity effects

This section presents an overview of connectivity/reconstruction effects in Dutch restrictive RCs, i.e. connectivity effects between (material inside) the RC head and the RC internal gap.[61,62] If possible, the connectivity effects are illustrated in both short and long-distance RCs, and in long-distance RCs for the Standard Dutch variant (*die-dat*) as well as for the doubling variant (identical doubling with pronoun *die*: *die-die*). There is not any difference in connectivity effects between these two variants.[63,64] First, in sections 3.6.1 and 3.6.2, I present connectivity effects and the lack of connectivity effects ('anti-connectivity/reconstruction') in Dutch RCs respectively. Then I discuss each type of connectivity effect separately: idiom connectivity (section 3.6.3), scope connectivity (section 3.6.4), the low construal of adjectival modifiers (section 3.6.5), Principle A connectivity (section 3.6.6), and variable binding connec-

[61] In all examples in this section, I indicate the gap position inside the RC (as _), but I leave undetermined what moved from that position: only the relative pronoun/operator (as in a HEA) or the RC head *plus* the relative pronoun/operator (as in a HIA).

[62] In this section, I will not be concerned with connectivity effects between material inside the *operator phrase* (instead of material inside the *RC head*) and the RC internal gap, an example of which is provided in (i) (Condition C violation). The reason for this is twofold. First, connectivity effects between material inside the operator phrase and the RC internal gap are hard to test systematically (it is hard to come up with natural examples for all types of connectivity effects). Second, I am primarily concerned with the question of whether or not there is movement of the RC head (HEA vs. HIAs). Assuming an account of connectivity effects in terms of syntactic reconstruction (movement), this question is immediately related to the presence or absence of connectivity effects between material inside the *RC head* and the gap inside the RC. Connectivity effects between material inside the *operator phrase* and the gap inside the RC are less interesting from this point of view, as both the HEA and HIAs assume the operator must have moved from a RC internal position to the left periphery. Besides, as I will show in this section, connectivity effects turn out not to be a proper diagnostic for movement anyway.

(i) * I respect $[_{RC\,head}$ any writer$]$ $[_{RC}$ $[_{Op}$ whose depiction of John$_i$$]$ he$_i$'ll object to$]$.
[Sauerland 1998:65]

[63] Many thanks to Hilda Koopman for providing me with the relevant doubling data.

[64] It should be mentioned that the grammaticality judgments regarding connectivity effects (especially in long-distance RCs) are often less clear than indicated in this section. This emphasizes again that connectivity effects are notoriously delicate, as a result of which the discussion of them is rather subtle and complex.

tivity (section 3.6.7). I show for each type the problems that a syntactic reconstruction approach encounters and present alternative analyses to syntactic reconstruction (if existent) for the different types of connectivity involved.

3.6.1 Connectivity effects in Dutch RCs

Dutch restrictive RCs show idiom connectivity, Principle A connectivity, variable binding connectivity, scope connectivity, and connectivity regarding the interpretation of adjectival modifiers, as illustrated in (264)-(268).[65]

(264) *idiom connectivity*

a. De [**streek**] die hij me _ **leverde**, riep om wraak.
the nasty joke RP he me delivered cried for revenge
'The nasty joke he pulled on me cried for revenge.'
[de Vries 2002:78]

b. De [**streek**] die ik denk dat/die hij me _ **leverde**, riep om wraak.
the nasty joke RP I think that/RP he me delivered cried for revenge

(265) *Principle A connectivity*

a. die [rare verhalen over **zichzelf**$_i$] die **Paul**$_i$ gisteren _ te horen kreeg
those weird stories about SE-SELF RP Paul yesterday to hear got
'those weird stories about himself that Paul heard yesterday'
[adapted from de Vries 2002:80]

b. die [rare verhalen over **zichzelf**$_i$] die jij dacht dat/die **Paul**$_i$ gisteren _ te horen kreeg
those weird stories about SE-SELF RP you thought that/RP Paul yesterday to hear got

(266) *variable binding connectivity*

a. de [ouders van **zijn**$_i$ geliefde] die **iedere man**$_i$ graag _ wil ontmoeten
the parents of his beloved RP every man gladly wants meet
'the parents of his beloved that every man gladly wants to meet'

[65] Principle B connectivity is left out because it is hard (if not impossible) to be adequately tested (cf. Bhatt 2002, Salzmann 2006:100, Sportiche 2006 a.o.). As there can be coreference between the pronoun and its antecedent without the pronoun being c-commanded by its antecedent, coreference does not imply (anti)reconstruction.

b. de [ouders van **zijn**$_i$ geliefde] die ik denk dat/die **iedere man**$_i$
the parents of his beloved RP I think that/RP every man
graag _ wil ontmoeten
gladly wants meet

(267) *scope connectivity*
de [**band**] die **iedere student** _ het beste vindt
the band RP every student the best finds
'the band that every student likes best'
[$\exists > \forall$; $\forall > \exists$, cf. Salzmann 2006:95 for German]

(268) *the low reading of adjectival modifiers*
de *eerste* roman die je *zei* dat/die Tolstoj *geschreven* heeft
the first novel RP you said that/RP Tolstoy written has
'the first novel that you said Tolstoy has written'
[cf. Bhatt 2002:57 for comparable English examples]

a. high reading: the first novel about which you *said* that Tolstoy had written it

b. low reading: you said that the first novel that Tolstoy had *written* is x

3.6.2 Anti-connectivity effects in Dutch RCs

The RC head cannot always be interpreted RC internally, but must sometimes be interpreted in the matrix clause: anti-connectivity/reconstruction. Examples of this are given in (269)-(271) which show the lack of connectivity effects for idiom interpretation, anaphor binding and Principle C respectively ('lack of Principle C effect', cf. also *supra*).[66]

[66]The fact that the pattern of (lack of) Principle C effects is more complicated than the core case in (271) suggests, is irrelevant for present purposes (but see Salzmann 2006 for a detailed overview). I only want to mention here argument/adjunct asymmetries as regards Principle C. It has often been mentioned that only R-expressions in *arguments*, but not R-expressions in *adjuncts*, bring about condition C effects (cf. Freidin 1986, Lebeaux 1988, 1990, Fox 1999b a.o.). This is illustrated for *wh*-movement by the contrast in (i)-(ii).

(i) * [Which claim [$_{ARGUMENT}$ that Mary had offended John$_i$]] did he$_i$ repeat _?

(ii) [Which claim [$_{ADJUNCT}$ that offended John$_i$]] did he$_i$ repeat _?

This asymmetry is usually accounted for by assuming that adjuncts (like the RC in (ii)) can be *merged late* (countercyclic merger, cf. Lebeaux 1988, 1990, Chomsky 1993 a.o.) – thereby accounting for the lack of Principle C effect in (ii) as the R-expression is not c-commanded by the pronoun at any point in the derivation. However, as this approach to the argument/adjunct asymmetry has been challenged on both empirical and theoretical grounds (cf. Heycock 1995, Lasnik 1998, Fischer 2002 a.o.), I will not be concerned with the apparent argument/adjunct distinction in the context of Condition C effects.

(269) a. Hij **leverde** een [**streek**] die _ om wraak riep.
he delivered a nasty joke RP for revenge cried
'He pulled a nasty joke that cried for revenge.'

b. Hij **leverde** een [**streek**] die jij vindt dat/die _ om wraak riep.
he delivered a nasty joke RP you find that/RP for revenge cried

(270) a. **Paul**$_i$ haat die [rare verhalen over **zichzelf**$_i$] die Marie _ steeds weer vertelt.
Paul hates those weird stories about SE-SELF RP Marie time and again tells
'Paul hates those weird stories about himself that Marie tells time and again.'

b. **Paul**$_i$ haat die [rare verhalen over **zichzelf**$_i$] die ik geloof dat/die Marie _ steeds weer vertelt.
Paul hates those weird stories about SE-SELF RP I believe that/RP Marie time and again tells

(271) a. de [vriend van **Jan**$_i$] die **hij**$_i$ _ zo aardig vindt
the friend of Jan RP he so nice finds
'the friend of Jan that he thinks is so nice'

b. de [vriend van **Jan**$_i$] die je denkt dat/die **hij**$_i$ _ zo aardig vindt
the friend of Jan RP you think that/RP he so nice finds

Finally, Negative Polarity Item (NPI) connectivity can be absent in Dutch RCs (cf. Citko 2001 for English). The Immediate Scope Constraint (Linebarger 1980) requires an NPI to be in the immediate scope of the negation operator at LF, i.e. no other scope-bearing element can intervene between the NPI and its licensor. The workings of this constraint are illustrated in (272b): the quantifier *iedere* 'every' intervenes between the NPI *een rooie cent* 'a red cent' and its licensor *niemand* 'nobody', resulting in ungrammaticality. However, in RCs we do not find such intervention effects, as illustrated in (273), which suggests that the RC head containing the NPI does not reconstruct.[67]

[67]It is unclear whether or not these sentences indeed show that the RC head does not reconstruct for NPI licensing. It has been argued in the literature that NPI licensing is sensitive to *surface structure* (see e.g. Sternefeld 2001). If true, the attested reconstruction behavior of NPIs (in RCs) cannot be accounted for by means of (the absence or presence of) *syntactic reconstruction* (movement).

(272) a. *Niemand* gaf die liefdadigheidsinstelling *een rooie cent*.
nobody gave that charity a red cent
'Nobody gave a red cent to that charity.'

b. * *Niemand* gaf ***iedere*** liefdadigheidsinstelling *een rooie cent*.
nobody gave every charity a red cent

(273) a. *Niemand* vond een [foto van *ook maar één* meisje].
nobody found a picture of any girl

b. *Niemand* vond een [foto van *ook maar één* meisje] [die
nobody found a picture of any girl RP
iedereen _ mooi vond].
everybody beautiful found

c. *Niemand* vond een [foto van *ook maar één* meisje] [die hij
nobody found a picture of any girl RP he
denkt [dat/die ***iedereen*** _ mooi vond]].
thinks that/RP everybody beautiful found

Under the assumption that connectivity effects signal movement (but see *infra*), the anti-connectivity/reconstruction effects in (269)-(273) suggest a HEA of RCs, whereas the connectivity effects in (264)-(268) suggest the opposite.[68] The aim of the following sections is not to give a full-fledged analysis of the attested connectivity effects, but rather to show that the presence of connectivity effects does not necessarily signal movement.

3.6.3 Idiom connectivity

The fact that (some) idioms can be relativized has always been taken as a strong argument in favor of HIAs of RCs, because in order to get the idiomatic interpretation, the RC head needs to reconstruct to its base position at LF due to the *adjacency requirement* on idiom interpretation. However, Lasnik and Fiengo (1974:541) observe that, in contrast to the famous and often used VP idiom *making headway* in (274), some object NPs of VP idioms can in

[68] As already pointed out in section 3.2.3, the anti-connectivity/reconstruction effects in (269)-(273) are especially problematic for a *raising* analysis of RCs, because under such an analysis the lowest copy will be interpreted by default, predicting reconstruction effects to arise across the board. The anti-connectivity/reconstruction effects are less problematic for a *matching* analysis, because within a MA there is always the option of interpreting the RC external head instead of a copy of the RC internal head (i.e. the lower copy can delete at LF when its content is *recoverable* from the external head, cf. Munn 1994, Citko 2001). More generally, the lack of connectivity effects might not be really problematic for HIAs as they can potentially be accounted for by assuming that in RCs (in contrast to *wh*-Qs) syntactic reconstruction is *not* the default, i.e. there is only syntactic reconstruction when it is forced for some reason (e.g. for idiom interpretation (264) or variable binding (266)).

There is, however, a caveat to this: under a HIA of RCs, the RC head must be interpreted in its base position for *theta reasons*. So, it cannot be simply assumed that the head is only interpreted in a higher position. Notice that no such problem arises for the HEA of RCs, because the operator/relative pronoun saturates the argument position of the verb.

fact not relativize, as illustrated in (275). Under a HEA the impossibility of an idiomatic interpretation under relativization follows, but under a HIA these examples are more problematic, i.e. a HIA incorrectly predicts the sentences in (275) to be grammatical.[69] When we now consider the grammaticality of the examples in (276) in which the idiom has been passivized, the ungrammaticality of the sentences in (275) becomes even more puzzling: a HIA predicts the same outcome in both types of sentences as both sentence types are derived by movement of the object NP. A HEA on the other hand actually predicts this pattern: only in the case of passivization have the NP and the verb originated as one constituent and can they get the idiomatic interpretation. In RCs, however, the RC head and the verb have never been a constituent and the idiomatic interpretation under relativization is correctly predicted to be out.[70,71] Another example of this pattern is given in (277); when the NP complement of the idiom *spill the beans* is relativized, as in (277b), it cannot get the idiomatic interpretation but it can get a literal interpretation (hence the symbol # in front of the sentence).

(274) The headway that we made was sufficient.

(275) a. * The heed that we paid to that warning was slight.
b. * The attention that we paid to the lecture was careful.
[Lasnik and Fiengo 1974:541]

(276) a. Heed was paid to our warning.
b. Attention was paid to our problems.
[Lasnik and Fiengo 1974:541]

(277) a. The beans appeared to be spilled when he opened his mouth.
b. # The beans that Joe spilled caused us a lot of trouble.
[Horn 2003:262]

[69]The grammaticality judgments in (274)-(276) are taken from Lasnik and Fiengo (1974), but not all people I consulted agree on the status of these sentences.

[70]In most cases an idiom can relativize, which might at first sight be problematic for the HEA. However, the important thing to note here is the contrast in grammaticality between the RCs and the passives, which is predicted under a HEA. Further research into the properties of idioms that can and idioms that cannot relativize (and their corresponding passivization patterns) should shed more light on the issue.

[71]De Vries (2002) distinguishes between 'real' idioms, which are established holistically and of which the meaning cannot be determined by the literal meaning of the component parts (*semantic idioms*), and *collocations*, most of which involve a semantically bleached/light verb, like *take a dive/swim/shower*, *make progress/headway*. Only the latter, but not the former type of collocation can *in principle* be split across a relative construction. According to de Vries (2002:79) this is obvious "since it is not possible to relate two meanings at once to the head noun: an idiomatic one in the relative and a literal (or 'decomposed') one in the matrix." It is important to note that we are not dealing with real semantic idioms in the case of (275), and that strings like *pay slight heed to* and *pay careful attention to* are perfectly fine (with the idiomatic interpretation). The sentences in (275) thus cannot be ruled out on other grounds than relativization of the idiom (see main text).

This line of reasoning – HIAs cannot explain the contrast between (275) and (276), whereas a HEA can – relies on the assumption that the presence or absence of connectivity effects (here for idiom interpretation) reliably diagnoses the presence or absence of movement. Recently, Webelhuth (2011) (in part based on Wasow et al. 1984, Nunberg et al. 1994, Jackendoff 1997, Horn 2003 a.o.) has proposed a *lexico-syntactic* analysis of idiom interpretation (across different construction types) that does not rely at all on syntactic reconstruction. By postulating specific lexical entries and constraints, he accounts for idiom interpretation in RCs (and other construction types) without literal reconstruction. Whether or not a particular idiom can appear in RCs, is dependent on its lexical entry. If an idiom has a phrasal lexical entry (i.e. it is thematically non-compositional) it cannot occur in RCs (e.g. *spill the beans*, *kick the bucket*). If, on the other hand, the lexical entry of (part of) an idiom is a word (i.e. the idiom is thematically compositional: the NP complement may occur in its idiomatic sense with other verbs than the verb in the idiom), it may occur in RCs, e.g. *make headway* as in (274), or *pull strings*, as illustrated in (278). In that case the relative pronoun (or operator) can satisfy the requirements on the nominal part of the idiom.[72]

(278) a. Strings seem to be pulled every time he applies for a promotion.
b. We were surprised at the strings that were pulled to get Joe's promotion. [Horn 2003:261]

3.6.4 Scope connectivity

A good alternative to *syntactic* reconstruction to account for scope interpretation is *semantic* reconstruction, i.e. the interpretation of scope inversion by means of semantic methods, e.g. Cresti (1995), Rullmann (1995), Sternefeld (2001). An often mentioned argument in favor of *syntactic* reconstruction and against *semantic* approaches to reconstruction is the unification for Binding Reconstruction and Scope Reconstruction (e.g. Romero 1997, Fox 1999a), as illustrated in (279). These examples show that whenever there is reconstruction for binding (279a), there is reconstruction for scope (the existential can get narrow scope), but when there is no reconstruction for binding (279b), the narrow scope interpretation of the existential is not available.

(279) a. [A student of **his**$_i$]$_1$ seems to **David**$_i$ $\_1$ to be at the party. [*seem*>∃; ∃>*seem*]
b. [A student of **David**$_i$'s]$_1$ seems to **him**$_i$ $\_1$ to be at the party. [**seem*>∃; ∃>*seem*, Fox 1999a:170]

Cecchetto (2001) shows that Italian *Clitic Left Dislocation* (CLLD) with a dislocated PP provides evidence against the unification of Binding Reconstruction

[72] Future research is required to see if idiom connectivity can get a (partly) semantic account (Webelhuth 2011:59).

and Scope Reconstruction. The sentence in (280) shows that there is Binding Reconstruction of the dislocated PP (i.e. *pro* cannot refer to *Gianni* due to a Principle C violation) but that same PP cannot reconstruct for Scope (i.e. there is no corresponding narrow scope reading of the existential). Fox's (1999a) argument in favor of *syntactic* and against *semantic* approaches to reconstruction thus does not stand. Consequently, scope reconstruction cannot provide a case for either *syntactic* or *semantic* reconstruction.

(280) In una casa di Gianni$_i$ pro$_j$ ci ha ospitato ogni ragazzo.
in a house of Gianni (he) there has hosted every boy
'In a house of Gianni every boy was hosted.'
[$\exists > \forall$; *$\forall > \exists$, Cecchetto 2001:2]

3.6.5 The low reading of adjectival modifiers

Bhatt (2002) argues that syntactic reconstruction of the RC head and its modifier(s) accounts for the possible *low* reading of the adjectival modifier, i.e. superlative adjectives, ordinals, nominal *only* and numerals. However, Heycock (2005) shows that this account of the low reading of the adjectival modifier "overgenerates massively" (pg. 380). First – as also noted by Bhatt (2002:73) himself in a footnote – not all adjectival modifiers allow for a low reading, and more importantly, the low readings of adjectival modifiers are not always generated by virtue of syntactic reconstruction, i.e. similar readings may arise in the absence of a RC.[73] This is illustrated for an *evaluative* adjective in (281)-(282) – notice that *wonderful* requires a 'scare quote' intonation. More specifically, the judgment in (281) that the books are wonderful can be ascribed to the speaker or to Siouxsie (*low* reading), but for the apparent low reading no RC is required, as illustrated in (282). In short, whatever accounts for the low reading of the evaluative adjective in (282), it is most likely not syntactic reconstruction.[74]

(281) the wonderful books that Siouxsie said that Lydia has written
[Bhatt 2002:73]

(282) Siouxsie was always going on about the books that Lydia has written. But I've read those wonderful books and they're complete rubbish.
[Heycock 2005:362]

Furthermore, Bhatt (2002) takes the observation that the low reading of the adjectival modifier is blocked by intervening negation, as illustrated in (283)

[73] It is furthermore unclear if the adjective originates within the RC in the first place (cf. section 3.5.1).

[74] Notice that in principle (282) could be derived from *But I've read those wonderful books that Siouxsie said that Lydia has written and they're complete rubbish* by ellipsis of the RC: *But I've read those wonderful books ~~that Siouxsie said that Lydia has written~~ and they're complete rubbish*. Under such a scenario, the low reading of the evaluative adjective can in fact be accounted for by means of syntactic reconstruction (by assuming a HIA of RCs).

for the adjectives *first* and *longest* respectively, to provide an argument in favor of his analysis of the low reading of adjectival modifiers: negation blocks reconstruction of the adjectival modifier.

(283) a. This is the first book that John didn't say that Antonia wrote.
b. This is the longest book that John didn't say that Antonia wrote.

Interestingly, however, Heycock (2005) shows that this intervention effect only holds for the low reading of adjectival modifiers but cannot be reproduced for e.g. binding for Principle A. To illustrate this, consider the sentence in (284). Whereas the modifier *only* cannot reconstruct due to the presence of negation, the RC head containing the reflexive *himself* must reconstruct for Principle A. These conflicting reconstruction requirements cannot be accounted for by assuming that the low reading of adjectival modifiers as well as binding for Principle A are derived by syntactic reconstruction. At the very least, this *dissociation* of different connectivity effects (cf. footnote 17) casts doubt on the reliability of taking (all) connectivity effects as a diagnostic for movement in general,[75] and on the analysis of the low reading of adjectival modifiers in terms of syntactic reconstruction in specific. Heycock (2005) argues that the low reading of an adjectival modifier in the RC head does not come about by syntactic reconstruction (but rather by interpreting the negation in the entailment with lower scope; see Heycock 2005 for details).

(284) This is the only picture of himself$_i$ that Mary didn't think John$_i$ should show to his mother. [Heycock 2005:366]

The low reading of (some) adjectival modifiers thus cannot be used as an argument in favor or against either the HEA or HIAs of RCs. Put differently, the low reading of adjectival modifiers provides inconclusive evidence to support any analysis of RCs, as it is unclear whether or not syntactic reconstruction is involved in the analysis of the phenomenon.

3.6.6 Principle A connectivity

Constructions in which an anaphor contained in the RC head is bound by an antecedent inside the RC have been taken as robust evidence for syntactic reconstruction of the RC head and thus for HIAs of RCs. However, in RCs there can also be binding for Principle A without a copy being present in the c-command domain of the antecedent (cf. Cecchetto 2005), as illustrated in (285).[76]

[75] Similarly, Cecchetto (2005) argues that only a uniform pattern of connectivity effects is a reliable diagnostic for movement, but when there is a dissociation of different connectivity effects, they should be accounted for by a semantic mechanism, not by syntactic reconstruction.

[76] As pointed out by Cecchetto (2005) a.o., testing syntactic reconstruction for Principle A with structures in which the RC head contains a transitive noun like *picture* is not reliable,

(285) % de [mislukking van **zichzelf**$_i$] die _ **hem**$_i$ beroemd heeft gemaakt
the failure of SE-SELF RP him famous had made
[164/380=43%, MPQ2-B data]

More generally, binding for Principle A without syntactic reconstruction is attested in other constructions as well, as illustrated for left dislocation in (286) and for topicalization in (287).[77] Both sets of examples show that syntactic reconstruction of the fronted constituent would lead to selectional problems, thus suggesting that this constituent is base-generated in the left periphery instead of being moved there (cf. Zwart 1993:109ff. for similar examples without bound elements in the fronted constituent).

(286) a. [Elkaar$_i$ helpen] dat doen ze$_i$ hier niet _.
each other help that do they here not
'Help each other, they don't do that here.'

b. * Ze$_i$ doen hier niet [elkaar$_i$ helpen].
they do here not each other help [Hoekstra 1999:65]

---

because transitive nouns may have an internal subject PRO that can bind the anaphor inside the RC head. More specifically, although a sentence like (ia) is often used to argue in favor of syntactic reconstruction of the RC head for binding purposes, this sentence might very well have a structure as given in (ib). In this structure an implicit subject PRO binds the anaphor (*himself*), and the PRO itself is controlled by *John*. As there is no c-command requirement on this backward pronominalization configuration, no syntactic reconstruction of the RC head is necessary for the anaphor to be properly bound. The lack of the need for c-command is illustrated by the Italian example in (iia). The acceptability of this example cannot be explained by syntactic reconstruction of the RC head into the gap position, because this gap position is not c-commanded by the alleged antecedent of the anaphor: *Gianni*. The grammaticality of (ii) can thus only be explained by assuming that there is an implicit PRO in the subject position of the NP that binds the anaphor, and that this PRO can be controlled by *Gianni* without there being c-command. This is illustrated in (iib)

(i) a. the [picture of himself$_i$] that John$_i$ likes _ most
b. the [$_{NP}$ PRO$_i$ picture of himself$_i$] that John$_i$ likes _ most
[Cecchetto 2005:16]

(ii) a. la descrizione di se stesso che _ aiuterebbe Gianni a passare l'esame
the description of himself that would help Gianni to pass the exam
b. la [$_{NP}$ PRO$_i$ descrizione di se stesso$_i$] che _ aiuterebbe Gianni$_i$ a passare l'esame
[Cecchetto 2005:17]

In sum, to explain the lack of Principle A effect in (i) and (ii), it is not necessary (and not even possible in the case of (ii)) to assume that a copy of the RC head is present inside the RC. Rather, an implicit subject PRO can (and must in the case of (ii)) bind the anaphor. Sentences like (i) and (ii) thus do not provide evidence in favor of HIAs of RCs. Following Cecchetto (2005:16ff.), I controlled for this complication regarding the testing of Principle A connectivity in RCs, by using an *unaccusative* noun as the RC head in (285). Only in that case can we be sure that there is no implicit PRO binding the anaphor inside the RC head, as there is simply no slot for an NP internal subject PRO. The noun *mislukking* 'failure' is derived from the intransitive unaccusative verb *mislukken* 'to fail' (*mislukken* is unaccusative according to the standard tests of unaccusativity, e.g. it takes the auxiliary *zijn* 'be', and it cannot occur in the impersonal passive construction), as a result of which it inherits its thematic grid from this verb.

[77] Thanks to Marcel den Dikken for drawing my attention to examples like these.

(287) a. [Elkaar$_i$ kussen] hebben zij$_i$ nooit geprobeerd.
each other kiss have they never tried
[322/380=85%]

b. ?* Zij$_i$ hebben nooit [elkaar$_i$ kussen] geprobeerd.
they have never each other kiss tried
[10/380=3%, MPQ2-B data]

3.6.7 Variable binding connectivity

The examples in (288) show that there can be variable binding in the absence of c-command (albeit only marginally). That is to say, in order for the quantifier *every man* in the examples in (288) to c-command and bind the variable *him*, it would have to move out of the RC, thereby violating the local character of quantifier raising (QR) and the Complex NP Constraint (CNPC, Ross 1967). It thus seems impossible that the binding relations in (288) are established by means of syntactic reconstruction.

(288) a. The woman [$_{RC}$ that **every man**$_i$ invited _] thanked **him**$_i$.
[Salzmann 2006:55]

b. ?? The woman [$_{RC}$ **every man**$_i$ invited _ to the party] came without **him**$_i$. [Sharvit 1999b:448]

More generally, variable binding without syntactic reconstruction occurs in other constructions as well, e.g. in left dislocation structures (289)-(290), specificational pseudoclefts (292), *tough*-movement constructions (293), and identity sentences (295).

On the basis of the Dutch left dislocation structure in (289), van Craenenbroeck (2004) shows the need for reconstruction without movement. More specifically, in order for the pronoun *zijn* 'his' in the left dislocated constituent to be bound by the quantifier *iedere taalkundige* 'every linguist' it would need to reconstruct. However, it crucially cannot reconstruct, because the left dislocated element is a PP whereas the reconstruction position is not.[78]

(289) Naar **zijn**$_i$ promotie, daar kijkt **iedere taalkundige**$_i$ naar _ uit.
to his defense there looks every linguist to out
'Every linguist looks forward to his defense.'
[van Craenenbroeck 2004:48]

[78]Similarly, Cinque (1990:chapter 2) argues at length that Clitic Left Dislocation (CLLD) in Romance should be analyzed by base-generation of the left dislocated phrase in the left periphery, even though there can be obligatory connectivity between (material in) the left dislocated phrase and the IP-internal argument position, as illustrated in (i) for binding connectivity.

(i) A se stessa, Maria non ci pensa
of herself, Maria not-there-thinks
'Maria doesn't think of herself.' [Italian, Cinque 1990:59]

Similarly, Guilliot and Malkawi (2007) observe that – given the uncontroversial assumption that movement out of strong islands is blocked – the fact that we do find variable binding connectivity effects in adjunct islands (with resumption), as illustrated for the French left dislocation structure in (290), shows that connectivity effects cannot be a proper diagnostic for movement.[79]

(290) [La photo de **sa**$_i$ classe]$_k$, tu es fâché [$_{ADJUNCT}$ parce que **chaque prof**$_i$ l$_k$'a déchirée].
'The picture of his class, you're furious because each teacher tore it.'
[Guilliot and Malkawi 2007:118]

Other examples of connectivity effects that cannot solely be accounted for by means of syntactic reconstruction are illustrated by specificational pseudoclefts (cf. Sharvit 1999a, Cecchetto 2001 a.o.). Specificational pseudoclefts, an example of which is given in (291), are a type of cleft sentence in which a *wh*-phrase – *wat Jan kocht* 'what Jan bought' in (291) – is equated with a constituent that corresponds to the gap inside the *wh*-phrase – *een boek* 'a book' in (291).

(291) [Wat Jan kocht _] was een boek.
what Jan bought was a book

Although it has been proposed in the literature that the pivot constituent – *een boek* 'a book' in (291) – and the gap inside the *wh*-phrase are related by syntactic movement, movement of the pivot from a position inside the *wh*-phrase to its surface position has several weird properties (cf. Cecchetto 2001 a.o.): it is a case of lowering movement, and it is movement from a *wh*-island (but see Bošković 1997 for an LF movement account). If there is no movement relation

[79] Ott (2011) proposes to analyze Contrastive Left Dislocation (CLD) – and dislocations more generally – in terms of the juxtaposition of two CPs *plus* IP ellipsis in one of the CPs (cf. the analysis of specificational pseudoclefts by den Dikken et al. 2000, footnote 80). This is illustrated in (i).

(i) *Contrastive Left Dislocation* (Ott 2011:13)
[$_{CP1}$ XP$_i$ ~~[$_{IP}$ ... t_i ...]~~] [$_{CP2}$ *d-pronoun*$_k$ ... t_k ...]
where XP = contrastive topic

Under this proposal the Dutch sentence in (289) receives an analysis as in (ii).

(ii) [$_{CP1}$ [naar *zijn* promotie]$_1$ ~~[$_{IP}$ kijkt *iedere taalkundige* _$_1$ uit]~~] [$_{CP2}$ daar$_2$ kijkt iedere taalkundige naar _$_2$ uit]

This analysis can account for the attested variable binding – i.e. the pronoun *zijn* 'his' in the CLDed PP gets bound by the universal quantifier *iedere taalkundige* 'every linguist' – by means of syntactic reconstruction: as the CLDed PP *naar zijn promotie* moves internally to its clause, variable binding can be computed within the elliptical CP$_1$. However, it is unclear how Ott's juxtaposition plus ellipsis analysis in its present form can account for the connectivity effects in sentences like (290), in which movement (of the CLDed phrase) is independently blocked. Additional assumptions are needed to account for such constructions. Put differently, even though the connectivity effect in (290) is most likely accounted for in terms of some sort of *ellipsis* (see also Guilliot and Malkawi 2007), it cannot be analyzed in terms of *movement*.

between the pivot and the gap inside the *wh*-phrase,[80] under the assumption that connectivity effects signal movement (syntactic reconstruction), we do not expect specificational pseudoclefts to show connectivity effects. However, as is well known, specificational pseudoclefts *do* show connectivity effects (cf. Higgins 1979, Sharvit 1999a, den Dikken 2006b amongst many others), as illustrated for variable binding connectivity in (292): the pivot constituent seems to be interpreted in the position of the gap inside the *wh*-phrase. The bound variable reading in (292) most likely also does not result from LF scoping the quantifier to a position in which it c-commands the pronoun, as this movement would violate the local character of QR and it would move the quantifier out of a *wh*-island (cf. Cecchetto 2001:8). We thus seem to have another case of connectivity effects that cannot be accounted for by syntactic reconstruction.[81]

(292) [Wat **elke generaal**$_i$ verdedigde _] was **zijn**$_i$ bataljon.
what every general defended was his battalion
[translated from example (29a) in Cecchetto 2001]

As pointed out to me by Marcel den Dikken, also *tough*-movement constructions seem to provide a case in point. Although we do find connectivity effects in *tough*-movement constructions, as illustrated for variable binding in (293),[82]

[80] An analysis of specificational pseudoclefts (SPCs) that does not assume movement of the pivot constituent from a position inside the *wh*-phrase, yet does account for the connectivity effects by means of syntactic reconstruction, is proposed by den Dikken, Meinunger, and Wilder (2000). They take the syntax of (a particular type of) SPCs to be on a par with the syntax of *Question-Answer pairs* (i), meaning that the *wh*-phrase in SPCs is a question and the pivot constituent is an IP that may be targeted by ellipsis (ii) (see Cecchetto 2001 for arguments against such an analysis of SPCs). Although this may account for the connectivity effects found in this particular type of SPC (iii), as den Dikken et al. (2000) point out themselves, not all SPCs can be analyzed this way, and for those SPCs it remains to be seen whether or not syntactic reconstruction can account for the attested connectivity effects.

(i) what did John buy? – [(he bought) some wine]

(ii) a. what John did/??bought was [$_{IP}$ he bought some wine]
b. what John bought was [$_{IP}$ ~~he bought~~ some wine]

(iii) wat *elke generaal*$_i$ verdedigde was [~~elke generaal$_i$ verdedigde~~ *zijn*$_i$ bataljon]
what every general defended was ~~every general defended~~ his batallion

[81] It was already noted by Chomsky (1981:346) that in sentences like (i) and (ii) "some sort of reconstruction seems necessary, though the trace position in which the matrix subject phrase is somehow interpreted is not the position from which it moved". Put differently, not all connectivity effects – here Principle A connectivity in a pseudocleft (i) and variable binding connectivity in an identity sentence (ii), cf. *infra* – can be accounted for in terms of syntactic reconstruction.

(i) pictures of **each other**$_i$ are what **they**$_i$ like to see

(ii) **his**$_i$ brother is the person whom **everyone**$_i$ admires most

[82] *Tough*-movement constructions show idiom connectivity as well, as illustrated in (i).

(i) *Headway* is difficult to *make* under such circumstances.

there is arguably no *syntactic* reconstruction in *tough*-movement constructions as the surface subject is not subject to the restrictions of the lower predicate (294): *to believe* does not select *for to* infinitives (Wilder 1991).[83]

(293) **His**$_i$ car is tough for **every man**$_i$ to have to part with.

(294) a. [For him to be top of the class] is hard to believe.
b. * It is hard to believe [for him to be top of the class].

Finally, as shown by Cecchetto (2005), in identity sentences (295a) – cf. the specificational sentences in (296) from Sharvit (1999a) – in contrast to canonical subject-predicate structures (295b), there can be variable binding in the absence of c-command. This shows that a HIA of RCs together with the Copy Theory of Movement (Chomsky 1993) is insufficient to account for the binding effects in (295), as we would predict reconstruction effects to arise (or not to arise) across the board.[84]

(295) a. [The one accident of **his**$_i$] [that **everyone**$_i$ remembers _] is the one that affected **him**$_i$ first.
b. * [The one accident of **his**$_i$] [that **everyone**$_i$ remembers _] affected **him**$_i$ first. [Cecchetto 2005:19]

(296) a. The woman [**no man**$_i$ listens to _] are **his**$_i$ wife and **his**$_i$ mother-in-law.
b. The book [that **every actor**$_i$ hopes to write _ some day] is **his**$_i$ autobiography. [Sharvit 1999a:300-301]

3.6.8 Interim summary

In this section, I tried to make plausible the claim that connectivity effects cannot always get a strictly syntactic account: some sort of semantic reconstruction mechanism is needed as well (see e.g. Cresti 1995, Rullmann 1995, Sharvit 1999a, Sternefeld 2001, Ruys 2011). At this point it is unclear whether semantic accounts of connectivity can adequately account for the whole range of connectivity effects in RCs (besides scope connectivity and variable binding connectivity; see Sternefeld 2001 for an attempt to extend the domain of semantic reconstruction). However, independently of whether or not the connectivity effects in RCs can be accounted for by semantics, the facts above

[83] But see Salzmann (2006:271ff.) for an ellipsis account of *tough*-movement constructions (cf. a matching analysis of RCs), and see Hartman (2009) and Hicks (2009) for recent analyses of *tough*-movement constructions according to which the matrix subject is derived by movement (instead of being base-generated in its surface position).

[84] Building on proposals by Jacobson (1994) and Sharvit (1999b) a.o., Cecchetto (2005) argues it is the semantics of identity sentences (and more particularly, the semantics of the functional reading) that gives rise to the bound variable interpretation, not actual syntactic binding.

show that *syntactic* reconstruction cannot be the only way to account for connectivity effects. Put differently, the presence of connectivity effects is not a proper diagnostic for movement and it thus cannot be used for distinguishing HIAs from the HEA of RCs. Needless to say, further research is necessary to see whether (a single mechanism of) semantic reconstruction can adequately account for the intricate patterns of connectivity effects in RCs.

3.7 Conclusion

In this chapter, I discussed the syntax of Dutch restrictive RCs. Starting from the observation that doubling of the relative pronoun in colloquial Dutch long-distance RCs (cf. chapter 2) is most easily compatible with a (traditional) Head External Analysis (HEA) of RCs, I proposed a specific implementation of the HEA. According to this analysis, the RC head is base-generated in the highest SpecCP position of the RC itself, whereas the relative pronoun or operator moves to a lower SpecCP position. (Strictly speaking this analysis is a Head Internal Analysis (HIA) because the RC head originates *inside* the RC. However, since the analysis does not involve movement/raising of the RC head to the left periphery – one of the most important features of HIAs of RCs – I consider it to be a HEA.) In addition to accounting for the doubling data, this analysis was shown to account for the whole range of variation in the left periphery of Dutch RCs as well, particularly doubly filled Comp facts.

I furthermore showed that although HIAs (*raising* or *matching*) have gained a lot of ground in recent years, choosing between a HEA and HIAs is certainly not a trivial matter. The HEA fares better in many respects: case mismatches between the RC head and the relative pronoun, selectional differences between the relative pronoun and its determiner or interrogative counterpart, and locality constraint violations. *Connectivity effects* between (material inside) the RC head and the RC internal gap, on the other hand, have always been taken to strongly argue *against* a HEA of RCs. That is, if connectivity effects in RCs are the result of the activation of a lower copy at LF (*syntactic reconstruction*), the presence of an (additional) RC head within the RC is required – as in a HIA – to account for such connectivity effects. I showed that this most prominent argument in favor of HIAs is not very well founded, as connectivity effects cannot always be accounted for by means of syntactic reconstruction. Put differently, reconstruction without movement or copies seems to be needed anyway, in RCs as well as in other configurations, e.g. left dislocation structures and specificational pseudoclefts. This strongly suggests that connectivity effects are not a foolproof diagnostic for movement, and that the presence or absence of connectivity effects in RCs thus provides inconclusive evidence to support any analysis of RCs.

CHAPTER 4

On relative pronouns and complementizers

Whereas the previous chapter dealt with the relative clause construction *as a whole*, this chapter takes a closer look at the elements involved *within* this particular construction. The chapter consists of two parts. The first part is devoted to the nature and status of pronouns that may appear in relative clauses (RCs), and on the other functions that such pronouns may have (*multipurpose pronouns*). The second part is mainly concerned with the relation between relative pronouns and complementizers.

4.1 Introduction

A significant amount of linguistic research has been directed towards the internal and external syntax of pronominal expressions. Whereas pronouns are traditionally taken to be the spell outs of *heads* (cf. Postal 1969, Abney 1987, Longobardi 1994 a.o.: pronouns are determiners occupying D^0), recently many scholars have convincingly argued for a more articulated structure of pronominal expressions (cf. Cardinaletti 1994, Ritter 1995, Noguchi 1997, Wiltschko 1998, Cardinaletti and Starke 1999, Koopman 1999, Déchaine and Wiltschko 2002, Harley and Ritter 2002, Rooryck 2003, van Koppen 2005, Neeleman and Szendröi 2007, Barbiers et al. 2009 a.o.). Based on both morphological and syntactic considerations, it has furthermore been claimed that the external syntax of pronouns (e.g. their distribution, their binding-theoretic status) is the result of their internal syntax and categorial status (cf. Cardinaletti 1994, Corver and Delfitto 1999, Koopman 1999, Déchaine and Wiltschko 2002 a.o.).

Many languages – among which Dutch (cf. chapter 2) – make use of what I will refer to as *multipurpose pronouns*: pronouns that can have more than

one function and may appear in more than one syntactic configuration. For example, the Dutch pronoun *die* may function as a demonstrative *and* as a relative pronoun, and the Dutch pronoun *wie* may function as an interrogative *and* as a relative pronoun (in colloquial Dutch). This is illustrated in (297) and (298) respectively. Recall furthermore that in colloquial Dutch both elements (*die* and *wie*) can introduce the lower clause of a long-distance *wh*-question (*wh*-Q) or a long-distance RC (see chapter 2 for details).

(297) a. Jij hebt **die** (man) niet gezien.
you have that.C man.C not seen
'You have not seen that (man).' *demonstrative*

b. Ik ken de man **die** jij gezien hebt.
I know the man RP you seen have
'I know the man you have seen.' *relative pronoun*

(298) a. **Wie** heb jij gezien?
who have you seen
'Who have you seen?' *interrogative pronoun*

b. % Ik ken de man **wie** jij gezien hebt.
I know the man who you seen have
'I know the man you have seen.' *relative pronoun*

The natural question arises as to what is the nature of these multipurpose pronouns and how can we account for their multiple functions. Are these different functions simply the result of *lexical ambiguity* or *accidental homophony*?

The option of having a multipurpose pronoun listed in the lexicon multiple times (*lexicalist approach*) is disregarded because it is inadequate on theoretical as well as on empirical grounds. First, taking a lexicalist approach to multipurpose pronouns is theoretically unattractive as it would result in the postulation of several construction specific lexical entries (cf. Wiltschko 1998). That is, we want to account for the different functions a single pronoun can have, without having to invoke construction specific statements – especially since one of the goals of generative grammar in general, and the *Minimalist Program* (Chomsky 1993 and later work) in particular, is to do away with construction specific and language specific rules as much as possible. Second, the lexicalist approach to multipurpose pronouns does not have any *explanatory value*. Under such an approach, the observations that (i) the existence of multipurpose pronouns is a systematic phenomenon in natural languages (cf. Bhat 2004 a.o.) and that (ii) the same multifunctionality is found in different languages, are a mere coincidence. Finally, the traditional claim that the different functions of multipurpose pronouns are simply the result of different lexical entries is insufficient, because it cannot insightfully account for the empirical observation that the meaning and function of a multipurpose pronoun is (partially) determined by the syntactic configuration it occurs in (cf. Postma 1994, Cardinaletti and Starke 1999,

and Koopman 1999, Cheng 2001 a.o.).[1]

These observations lead to the hypothesis that a multipurpose pronoun is a single lexical element that is morphosyntactically and semantically ***underspecified*** (cf. Rooryck 2003), as a result of which it may appear in more than one syntactic environment, and part of its meaning is determined contextually or configurationally (cf. Postma 1994).[2] This hypothesis, which forms the foundation of this chapter, is formulated in (299).

(299) hypothesis
a multipurpose pronoun has a single underspecified lexical entry

I thus pursue a strong *anti-homophony* hypothesis: rather than treating the different instances of multipurpose pronouns as distinct, I assume that the meaning and function of a multipurpose pronoun is the result of its morphosyntactic feature specification in interaction with its syntactic environment. The first part of this chapter focuses from this perspective in detail on the nature and form of (multipurpose) relative pronouns in Dutch.

In addition to appearing in several configurations as a *pronoun*, some multipurpose pronouns may function as a *complementizer* as well. That is to say, in many languages the finite declarative complementizer is identical in form to a pronominal element, e.g. demonstratives in West-Germanic (e.g. Dutch *dat*, English *that*, German *dass*), and *wh*-phrases in Romance (e.g. Latin *quod*, French *que*, Spanish *que*, Italian *che*).[3] This is illustrated in (300) for the Dutch element *dat* 'that'.

(300) a. Jij hebt **dat** (meisje) niet gezien.
you have that.N girl.N not seen
'You have not seen that (girl).' *demonstrative*

b. Ik ken het meisje **dat** jij gezien hebt.
I know the girl RP you seen have
'I know the girl you have seen.' *relative pronoun*

[1] Although there thus seem to be good arguments to discard a lexicalist approach to multipurpose pronouns, the same does not seem to hold for pure lexical categories like nouns. That is to say, the observation that a single noun may have multiple unrelated meanings – e.g. the Dutch noun *vorst* can mean frost or ruler – is most likely best accounted for by postulating multiple lexical entries for the noun.

[2] An alternative approach is the nanosyntactic approach to multipurpose pronouns, according to which a multipurpose pronoun is a single lexical element that is morphosyntactically ***overspecified***, as a result of which it may appear in more than one syntactic environment (in accordance with the *Superset Principle*). See section 2.5.4 for details.

[3] In many languages – among which certain languages from the Kwa family (cf. Aboh 2004) – *that*-type complementizers arguably developed from the verb *say*. My analysis of the pronominal-like complementizers in Germanic (and Romance) will have nothing to say about these verbal-like complementizers. However, it is worth mentioning here that Aboh (2011) puts forward a particularly interesting generalization. On the basis of a small typological study of 16 languages (plus creoles), he makes the observation that languages that do not have a pronominal-like complementizer, tend to lack articles in general. This makes the link between complementizers and articles/pronouns even stronger.

c. Ik denk **dat** jij het meisje gezien hebt.
I think that you the girl seen have
'I think that you have seen the girl.' *complementizer*

The second part of this chapter is devoted to investigating the relation between (relative) pronouns and complementizers. I will argue that although the Dutch complementizer *dat* and the demonstrative/relative pronoun *dat* are not the same lexical item, they are diachronically related in a way that is compatible with the hypothesis in (299).

This chapter is organized as follows. In section 4.2, I discuss the nature of (multipurpose) relative pronouns in Dutch, and the other functions that such pronouns may have. In section 4.3, I offer some reasoned speculation on the nature of complementizers and their relation with pronouns (drawing from recent literature). Section 4.4 presents arguments against a recent analysis by Kayne (2008, 2010) amongst others, who argues that (sentential) complementation is relativization and that complementizers are relative pronouns. The final section of this chapter challenges this claim further by providing a case study that illustrates the need to distinguish complementizers from relative pronouns, namely special *die* in dialectal Dutch.

4.2 On the nature of relative pronouns

A restrictive RC modifies an antecedent: the RC head. A relative pronoun or operator in a restrictive RC mediates the relation between the RC and the RC head. More specifically – as laid out in chapter 3 – movement of a relative pronoun or operator to the left periphery turns the proposition into a predicate (*lambda/predicate abstraction*). The RC predicate is related to the RC head by means of *Predicate Modification* (Heim and Kratzer 1998:95), which semantically amounts to *set intersection*. Relative pronouns are thus nothing more than *operators* that turn the clause into a predicate (by movement to the left periphery: '*wh*-movement') and connect the RC to the RC head, i.e. relative pronouns are *referential* or *anaphoric* in the sense that they indicate "coreference with a concept that has been expressed elsewhere in the sentence" (Bhat 2004:230).

Languages that make use of relative pronouns – mostly Indo-European languages (de Vries 2002:173)[4] – generally do not have a single lexical item that is only used in RCs, but make use of relative pronouns that are borrowed from another pronoun paradigm, e.g. the demonstrative pronoun paradigm or the interrogative pronoun paradigm. Relative pronouns can thus often be used outside of RCs, namely as e.g. demonstratives or interrogatives. Assuming that we do not want to postulate a lexical feature [+relative pronoun] – contra the lexicalist hypothesis (cf. *supra*) – these observations merit the conclusion that

[4]See de Vries (2002) for a detailed overview of the relativization strategies in a sample of 172 languages worldwide.

there is no such thing as a relative pronoun (Wiltschko 1998).[5] But what in fact constitutes a good relative pronoun?

4.2.1 What is a relative pronoun?

Borrowing liberally from a proposal by Wiltschko (1998) regarding the nature of relative pronouns, I take relative pronouns to be DPs that contain an empty noun and have an operator. The operator is the driving force behind movement to the left periphery – resulting in an operator-variable chain – thereby turning the proposition into a predicate. The NP headed by the empty noun provides a *range* for the operator-variable chain (cf. also Longobardi 1994). The value of the range is provided by the RC head. The categorial status of relative pronouns is DP, because the gap inside the RC acts as a DP (cf. Borsley 1997, Bianchi 2000, Alexopoulou 2006 a.o., and see chapter 3 (footnote 8)). These three properties that a relative pronoun must possess are summarized in (301).

(301) demands on relative pronouns
  a. relative pronouns are DPs
  b. relative pronouns contain an empty NP
  c. relative pronouns contain an operator (in SpecDP)

Assuming that personal pronouns are not operators, but merely spell out ϕ-features – i.e. personal pronouns constitute PhiPs (in the sense of Déchaine and Wiltschko 2002) and thus lack a DP layer (cf. van Koppen 2005 a.o.) – personal pronouns are excluded from appearing as relative pronouns. As mentioned by Wiltschko (1998:149ff.), the claim that personal pronouns lack structure when compared to DPs is illustrated by languages like German, where the morphological form of personal pronouns is more or less contained in the morphological form of definite determiners (thereby abstracting away from phonological and/or spelling differences between the two forms). This is illustrated in table 4.1.[6] Even Dutch – in which the patterns are not as clear and transparent as in German – has some forms that seem to indicate that personal pronouns have less structure than DPs as well: *d-at* (3rd person neuter demonstrative pronoun/determiner) vs. *'t* (3rd person neuter personal pronoun; compare English *that* vs. *it*), and *d-ie* (common gender demonstrative pronoun/determiner) vs. *ie* (enclitic 3rd person masculine personal pronoun).

[5] If there is no lexical feature [+relative pronoun], the question immediately arises as to why there are languages that have pronouns that are specific to RCs (cf. de Vries 2002:395-396: (at least) 11 out of 52 languages that make use of relative pronouns (in postnominal relatives and correlatives) have *specialized* relative pronouns). I hope to return to the issue of the nature and status as well as the diachrony of such specialized relative pronouns in future research. Suffice it to say for now that the claim that there is no such thing as a relative pronoun holds true for all Germanic and all Romance languages in the typological survey in de Vries (2002).

[6] See Wiltschko (1998) for synchronic, diachronic and cross-linguistic support for this decomposition of pronouns (i.e. personal pronouns have less structure than DPs).

Table 4.1: German (Wiltschko 1998:149)

	definite determiners		personal pronouns	
SG	M	N	M	N
NOM	d-er	d-as	er	es
DAT	d-em	d-em	ihm	ihm
ACC	d-en	d-as	ihn	es

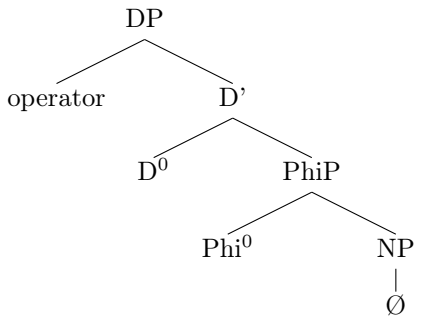
Just like personal pronouns, definite determiners do not contain an operator, as a result of which they cannot function as relative pronouns.[7]

Having eliminated both personal pronouns and definite determiners from acting as relative pronouns, we are left with the (A-bar) pronouns as illustrated in (302).[8] The syntactic structure of A-bar pronouns that I argued for in chapter 2 is repeated here as (303).

(302) a. *d*-pronouns (demonstratives): *die*, *dat* *deze*, *dit*
that.C that.N this.C this.N

b. *w*-pronouns (interrogatives): *wie*, *wat*, *welke*
who what which

(303) The structure of A-bar pronouns (= (58))



[7]Wiltschko (1998) argues that the operator-variable chain inside a RC needs a *range*. Only pronouns that contain an empty NP contain a range, and may thus function as relative pronouns. An empty NP can only be licensed by the presence of a D^0 and strong agreement. Since personal pronouns lack D^0, they lack an empty NP, as a result of which they are excluded from being used as relative pronouns. As for definite determiners in Dutch, Wiltschko (1998:161) argues that they cannot be used as relative pronouns, because they lack strong agreement: *de* 'the.C' and *het* 'the.N' (weak) versus *die* 'that.C' and *dat* 'that.N' (strong). Notice that *het* may in fact occur on its own (e.g. *het regent* 'it rains', *ik heb het gedaan* 'I have done it'), yet it cannot be used as a relative pronoun: **het meisje het ze geroepen hebben* 'the girl it they called have'. The observation that *het* cannot be used as relative pronoun is accounted for by the requirement that a relative pronoun needs an operator (301c).

[8]I will not be concerned with adjunct *wh*-phrases like *hoe* 'how', *waarom* 'why', and *wanneer* 'when', or with locative pronouns/adverbs *daar* 'there', *waar* 'where' ('oblique relative pronouns'), and *hier* 'here'. For aspects of the nature and distribution of locative pronouns see amongst others van Kampen (1997, 2007, 2010) and Rooryck (2003).

The *d*-pronouns *die* and *dat* – but crucially not *deze* and *dit*, cf. section 4.2.3 – and the *w*-pronouns *wie* and *wat* (and *welke* in formal Dutch, cf section 2.7.1), indeed occur in argument restrictive RCs. In chapter 2, I argued that both types of A-bar pronouns are equally suited to function as relative pronouns, the difference in their distribution being related to which features are spelled out: gender (common/neuter) or animacy (roughly human/non-human, but see the Individuation Hierarchy in (82) for a more fine-grained classification). I argued for the following (simplified) feature specifications of the relevant A-bar pronouns (underspecfied features are omitted).

(304) a. *wat*:
b. *dat*: [definite]
c. *wie*: [human]
d. *die*: [common], [definite]

Whereas in Standard Dutch the relative pronoun is required to spell out syntactic gender, in colloquial Dutch this grammatical distinction is less important and the relative pronoun may spell out semantic animacy instead. For example, in the case of a common gender human RC head like *man* 'man', Standard Dutch requires the *d*-pronoun *die* (that spells out the [common] feature), whereas colloquial Dutch allows the *w*-pronoun *wie* (that spells out the [human] feature) as well. This is illustrated in (305).

(305) a. de **man** ***die*** ze geroepen hebben
the man.C,HUMAN that.C they called have
'the man they have called'
b. % de **man** ***wie*** ze geroepen hebben
the man.C,HUMAN who.HUMAN they called have
'the man they have called'

Although in some cases and for some speakers *d*-pronouns and *w*-pronouns are interchangeable – in particular pronouns *die* and *wie*, cf. (305) and see also the doubling data involving these pronouns in chapter 2 – there are some distributional differences between them. These differences are the topic of the next section.

4.2.2 Differences between *d*-pronouns and *w*-pronouns

First, *w*-pronouns – in particular pronoun *wie* – can be the complement of a preposition, whereas *d*-pronouns (and *wat*) cannot, in a neutral environment. That is to say, a *d*-pronoun can only be used as the object of a preposition in informal speech, or when it is strongly emphasized (Haeseryn et al. 1997:307). In RCs, instead of the combination preposition plus *d*-pronoun, the combination preposition plus *w*-pronoun (*wie*) or the combination R-pronoun plus preposition is used (in case of a human antecedent this is possible only in informal speech), as illustrated in (306).

(306) a. * de man op die ik verliefd ben
the man on that.C I in love am

b. de man op wie ik verliefd ben
the man on who I in love am
'the man I am in love with'

c. % de man waarop ik verliefd ben
the man whereon I in love am
'the man I am in love with'

It was already pointed out by van Riemsdijk (1978:36ff.) that in the complement of a preposition, pronouns that are not inherently specified as [human] are replaced by an R-pronoun (which then inverts with the preposition by 'R-movement'). Similarly, Van Kampen (2007) puts forward the generalization in (307). Although it lacks explanatory power, this generalization accurately describes the behavior of A-bar pronouns with respect to prepositions. This is illustrated in (308). A-bar pronouns (and personal pronouns) that are not inherently specified as [human]/[animate] cannot occur in the complement of a preposition (308a), but rather invert with the preposition and become an R-pronoun (308b). Only *wie*, being inherently specified as [human] (and personal pronouns specified as [human]), can be the complement of a preposition (308c).[9]

(307) Pronouns that lack an inherent [+animate], lack the potential to realize an oblique [P pronoun]$_{PP}$

[(12) in van Kampen 2007]

(308) a. *op dat, *op die, *op wat, *op het
on that.N on that.C on what on it

b. daarop, waarop, erop
there-on where-on there-on

c. op wie, op hem/haar
on who on him/her

Second, *d*-pronouns cannot generally be used in possessive RC constructions, in contrast to German (e.g. Smits 1988).[10] Such constructions can only contain the *w*-pronoun *wie*, as illustrated in (309) – the use of *wier* in (309b) is formal (cf. ANS, Haeseryn et al. 1997:343).

[9]Unlike what I assume (cf. section 2.5.2.4 and see (308a)), van Riemsdijk (1978:37) takes *die* to be inherently specified as [human] and *op die* therefore to be a grammatical string of Dutch.

[10]The SAND data show that in 15 out of 243 locations in the Dutch speaking language area, pronoun *diens* may in fact occur in possessive RCs (although it is never the only possible possessive pronoun/strategy in those locations), cf. SAND1 data (Barbiers et al. 2005:89). A bit less than half of these locations are at the border with Germany.

(309) a. de jongen <wiens/*diens> moeder ik ken
the boy whose.GEN,M mother I know
'the boy whose mother I know'

b. de vrouw <wier/*dier> moeder ik ken
the woman whose.GEN,F mother I know
'the woman whose mother I know'

The generalization thus seems to be that whenever a relative pronoun is contained in a larger phrase, whether it be a prepositional phrase (306) or a possessive phrase (309), it will surface as the *w*-pronoun *wie* (or the R-pronoun *waar* plus preposition). Interestingly, de Vries (2002:chapter 8, 2006) argues that all types of possessive RCs in Dutch – as illustrated in (310) – are syntactically related in the sense that the relative pronoun/operator in all cases is embedded inside a PP. If this analysis of possessive RC constructions is on the right track, the observation that only *w*-pronoun *wie* is found in possessive RCs reduces to the generalization in (307), as there is always an (overt) preposition present.

(310) a. de man van wie ik de vader ken
the man of who I the father know
'the man whose father I know' [prepositional genitive]

b. de man wiens vader ik ken
the man whose father I know
'the man whose father I know' [morphological genitive]

c. de man wie z'n vader ik ken
the man who his father I know
'the man whose father I know'
[relative plus possessive pronoun, de Vries 2006:2]

Although at this point I do not have any deep insight to offer as to why (307) should hold, it is worthwhile mentioning that in older stages of Dutch, *d*-pronouns could appear in possessive RC constructions, e.g. *de man diens woord ick houde* 'the man whose word I kept' (van der Horst 2008:830). This form *diens* could be used to refer to masculine, feminine, and plural antecedents. The change from *die* to *wie* in the context of possessive RCs (and prepositional phrases) might be the reflection of the loss of the feature [human] on *die*, in line with the hypothesis in (307). As *die* in present-day Dutch is no longer *inherently specified* for [human] (cf. section 2.5.2.4) – although it is still *compatible* with a [human] antecedent – *die* is no longer found in the context of possessive RCs or prepositional phrases.

Finally, in free relative clauses (FRCs) in argument position (henceforth argument FRCs), *w*-pronouns are generally preferred over *d*-pronouns, as illustrated in (311a,b). These data are from MPQ1-A, cf. SAND1 data (Barbiers et al. 2005:90) for similar results in the Netherlands (predominantly in the Dutch speaking part of Belgium do informants sometimes allow only *die* in

FRCs).[11] Whereas about half of the speakers allow *die* as a relative pronoun in argument FRCs, *die* cannot (or only very marginally) occur as a relative pronoun in FRCs in adjunct position (henceforth adjunct FRCs), as illustrated in (311c,d).

(311) a. [Wie het weet] mag het zeggen.
who it knows may it say
'Who knows it, may say it.

b. % [Die het weet] mag het zeggen.
die it knows may it say
'Who knows it, may say it.'

c. [Wie het ook zei], hij gelooft het toch niet.
who it also said, he believes it surely not
'Whoever said it, he doesn't believe it anyway.'

d. ?* [Die het ook zei], hij gelooft het toch niet.
die it also said, he believes it surely not [MPQ1-A data]

The fact that a *d*-pronoun is not as easily used in an adjunct FRC (311d) as in an argument FRC (311b), suggests that adjunct FRCs differ from argument FRCs, in that the former resemble embedded *wh*-Qs (recall that *d*-pronouns are never allowed to introduce a *wh*-Q, cf. chapter 2). This opposition between argument and adjunct FRCs might very well be related to the interpretation of FRCs. It is well known that argument FRCs can be *definite* or *universal* (Larson 1987, Grosu and Landman 1998, de Vries 2004 a.o.): the relative pronoun in (311a,b) refers either to a specific person or to a person in general. The distribution of *d*-pronouns (inherently definite) and *w*-pronouns in argument FRCs (for speakers that allow both types of pronouns in these constructions) might thus be dependent on interpretation, i.e. the use of a *d*-pronoun in FRCs corresponds to the definite interpretation, whereas the use of a *w*-pronoun in FRCs corresponds to the universal interpretation. If adjunct FRCs always get a universal interpretation (which might be reflected by the fact that the distribution of the particle *ook* is restricted to adjunct FRCs), this would explain the fact that almost only *w*-pronouns are attested in adjunct FRCs. I leave this issue for further research.

This concludes the overview of the differences in distribution between *d*-pronouns and *w*-pronouns. The next two subsections discuss *d*-pronouns and *w*-pronouns in their use as relative pronouns, the most important questions being why only a subset of *d*-pronouns may function as relative pronouns, and

[11] In older stages of Dutch (i.e. early Middle Dutch) almost only *d*-pronouns were used in FRCs, whereas *w*-pronouns like *wie* and *wat* were almost never attested (cf. van der Horst 2008:377). The change from *d*-pronouns to *w*-pronouns in FRCs set in in late Middle Dutch (van der Horst 2008:377), but the change from *dat* to *wat* in FRCs started earlier than the change from *die* to *wie*. This difference between *dat/wat* and *die/wie* is reflected by the MPQ1-A data as well: the change from *d-* to *w*-pronouns seems to be completed only with *dat/wat*, i.e. *dat* only very marginally occurs as a relative pronoun in FRCs, whereas *die* as a relative pronoun in FRCs is accepted by about half of the speakers.

why *w*-pronouns can act as relative pronouns. Starting from the hypothesis in (299), the leading idea is that the more underspecified an element is, the more different syntactic environments it may appear in.

4.2.3 *d*-pronouns

Whereas the demonstrative pronouns *die* and *dat* are used as relative pronouns, the demonstrative pronouns *deze* and *dit* are not attested in RCs. Following the hypothesis in (299), the logical conclusion must be that the *proximal* demonstrative pronouns *deze* and *dit* are more specified than the *distal* demonstratives *die* and *dat*, as a consequence of which the distal but not the proximal demonstratives may be used in other configurations, like RCs. Rooryck (2003) claims that distal demonstratives *die* and *dat* are underspecified for *location* (i.e. distance to the *deictic center*), as indicated in (312), cf. Rigterink (2005); see Kayne (2008, 2010) for a more or less comparable proposal, which will be discussed in section 4.4. More specifically, distal demonstratives are identified by the absence of the feature proximal.[12]

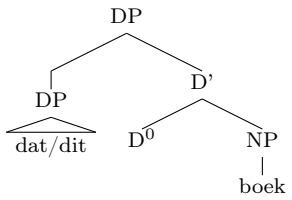
(312) a. *deze* (common)/ *dit* (neuter) 'this': [location: proximal]
b. *die* (common)/ *dat* (neuter) 'that': [location:]

As Rooryck (2003:11) puts it, when *die* and *dat* are used as demonstrative determiners/pronouns (i.e. without a RC antecedent),[13] they receive the distal interpretation by simply being part of the demonstrative paradigm to which the proximal demonstratives *deze* and *dit* belong, as a result of which *die* and *dat* can only take up the left over value in the paradigm, namely distal.[14] In

[12]Recall that I assume a representation of morphosyntactic features in terms of attribute-value structures, as discussed in section 2.5.2.1.

[13]I follow the standard assumption that demonstrative determiners occupy SpecDP (cf. Leu 2008:15 and references cited therein), as illustrated in (i) for the phrase *dat/dit boek* 'that.N/this.N book.N'.

(i)



[14]The definite determiners *de* 'the.C' and *het* 'the.N' in Dutch are not specified for location because they do not express space/place deixis in any syntactic environment. This might be formally implemented as follows: whereas definite determiners and demonstratives are all definite (in the sense that they refer to entities that are specific and identifiable), only demonstratives are further specified for location or distance to the deictic center (i.e. space/place deixis).

(i) a. [def : location : proximal] = *deze/dit* 'this' (and *hier* 'here')
b. [def : location :] = *die/dat* 'that' (and *daar* 'there')
c. [def :] = *de/het* 'the' (and *er* 'there')

terms of the lexicalization system I proposed in section 2.5.3, this can be formally implemented as illustrated in table 4.2. Something that is specified for location in syntax is best spelled out by a lexical item with the same specification for location: proximal in these cases (spelling out this syntactic feature representation by a lexical item that is underspecified for location is in principle possible, but will be blocked here by the more specified form). On the other hand, something that is underspecified for location in syntax can only be spelled out by a lexical item that is underspecified for location as well: distal in these cases (something underspecified in syntax cannot come out as more specified in morphology).

Table 4.2: Lexicalization of demonstratives

syntax	**lexicalization**	
[location: proximal]	[location: proximal]	e.g. *deze/dit*
[location: proximal]	[location:]	e.g. **die/dat* (blocked by *deze/dit*)
[location:]	[location:]	e.g. *die/dat*
[location:]	[location: proximal]	e.g. **deze/dit*

An argument in favor of the underspecification of *die* and *dat* for location comes from the pattern in (313), which shows that there is no feature clash in left dislocation constructions where a *distal* demonstrative is coreferent with a left dislocated *proximal* DP. If the distal demonstratives were to be specified as [distal], we would expect a feature clash between the distal demonstrative and the left dislocated DP which is specified as [proximal], *quod non.*[15]

(313) a. Deze jongen, die ken ik niet.
this.C boy.C that.C know I not
'This boy, I don't know.'

b. Dit meisje, dat ken ik niet.
this.N girl.N that.N know I not
'This girl, I don't know.' [adapted from Rooryck 2003:8]

[15]Something additional needs to be said about the fact that the sentences in (i) are ungrammatical. Whereas it is clear why distal demonstratives *may* occur in the left dislocation constructions in (313) (they are underspecified for location), it is unclear why they *must* occur in these constructions, and why proximal demonstratives cannot occur in these constructions. I propose that whereas the left dislocated constituent can be specified for location in syntax, the pronoun that moves to the left periphery is not specified for location in syntax. The left dislocated constituent and the pronoun thus only share their gender features. Only pronouns that are underspecified for location may spell out the pronoun in left dislocation structures: *distal* demonstratives.

(i) a. * Deze jongen, deze ken ik niet.
this.C boy.C this.C know I not

b. * Dit meisje, dit ken ik niet.
this.N girl.N this.N know I not

Furthermore, due to their underspecification for location, distal demonstratives can be used in neutral/non-deictic contexts (cf. Rigterink 2005), whereas proximal demonstratives are always forced to have a deictic interpretation in such environments (cf. Haeseryn et al. 1997:303: proximal *deze/dit* refer more explicitly/emphatically than distal *die/dat*). The examples in (314)-(316) illustrate this for evaluative contexts, familiar contexts and identification contexts respectively. The # sign in front of a pronoun indicates that the use of this pronoun in the given context is ungrammatical, unless the referent of the pronoun is physically present or present in the discourse.

(314) *evaluative contexts* (adapted from Rigterink 2005)

a. <*Die*/#*Deze*> idioot van een Jan heeft het weer gedaan.
that.C/this.C idiot of a Jan has it again done
'That idiot of a Jan has done it again.'

b. <*Dat*/#*Dit*> rotjoch van een Jan heeft het weer gedaan.
that.N/this.N bad boy of a Jan has it again done
'That bad boy of a Jan has done it again.'

c. <*Die*/#*Deze*> vreselijke/geweldige man!
that.C/this.C horrible/great man

d. <*Dat*/#*Dit*> vreselijke/geweldige boek!
that.N/this.N horrible/great book

(315) *familiar contexts* (adapted from Rigterink 2005)

a. en dan is er nog <*die*/#*deze*> kwestie van de mariniers
and then is there also that.C/this.C matter of the marines
'and then there is that matter relating to the marines'

b. en dan is er nog <*dat*/#*dit*> verhaal over die man
and then is there also that.N/this.N story about that man
'and then there is that story about the man'

(316) *identification* (cf. Dixon 2003:84)

a. het was <*die*/#*deze*> soort rijst waar Japanners dol op zijn
it was that.C/this.C sort rice where Japanese crazy on are
'it was that sort of rice that the Japanese love'

b. het is <*dat*/#*dit*> spul waar zij zo dol op zijn
it is that.N/this.N stuff where they so crazy on are
'it is that stuff that they love so much'

An indirect piece of evidence in favor of the claim that proximal demonstratives are specified for location is the presence of a gap in the paradigm, as illustrated in table 4.3. The absence of forms like *wier*, *wit* and *weze* is hard to account for when proximal demonstratives are underspecified for location, but can be easily accounted for under the assumption that proximal demonstratives are specified as proximal. More specifically, assuming that the specification for

location as proximal is incompatible with the question interpretation of the *w*-pronouns/adverb, accounts for the absence of forms like *wier*, i.e. a question interpretation (the speaker asks for the *unknown*) contradicts the demonstrative proximal meaning (close and therefore necessarily *known* to the speaker), cf. Rooryck (2003:11-12).[16]

Table 4.3: A gap in the paradigm (adapted from Rooryck 2003:11)

	d-adverb	*w*-adverb	*d*-pronoun	*w*-pronoun
proximal	hier (*dier)[17]	***wier**	dit/deze	***wit**/***weze**
distal	daar	waar	dat/die	wat/wie

In line with the hypothesis in (299), it is exactly the underspecification for location of distal demonstratives *die* and *dat* (i.e. the characterization of distal demonstratives as the absence of the feature proximal) that allows them to function as relative pronouns and topic *d*-pronouns as well. This is illustrated in (317) and (318) respectively.

(317) *relative clauses*

a. het boek <*dat*/**dit*> ik gelezen heb
the.N book.N that.N/this.N I read have
'the book I read'

b. de man <*die*/**deze*> ze geroepen hebben
the.C man.C that.C/this.C they called have
'the man they have called'

(318) *left dislocation*

a. het boek, <*dat*/**dit*> heb ik gelezen
the.N book.N that.N/this.C have I read

b. de man, <*die*/**deze*> hebben ze geroepen
the.C man.C that.C/this.C have they called

[16] Although according to Rooryck (2003) a question meaning and a demonstrative proximal meaning are inherently incompatible when combined within one and the same word, these meanings are in fact compatible when distributed over two words, e.g. *wie hier is zijn boek vergeten?* 'Who here has forgotten his book?' This observation weakens the (gap in the paradigm) argument in favor of the claim that proximal demonstratives are specified for location.

Interestingly, if I am right in claiming that an element can only be specified for location if it is specified for referentiality – cf. footnote 14 – the absence of forms like *wier* and *wit* follows straightforwardly. These forms are not specified as definite, and therefore cannot get a specification for location either, i.e. a single lexical item cannot be simultaneously underspecified for definiteness and specified for location. However, two separate lexical items can express the two meanings perfectly fine.

[17] I have no insight to offer as to why the proximal *d*-adverb features an *h*- instead of the *d*-morpheme. See Rooryck (2003:12) for some speculation on this matter.

Table 4.4 gives the feature specifications for the relevant *d*-pronouns (underspecified feature representations are omitted; see sections 2.5.2.4 and 2.5.2.5 for arguments justifying the feature specifications of *die* and *dat* respectively).[18] As table 4.4 shows, *dat* and *die* are the most underspecified pronouns in the *d*-pronoun paradigm, as a result of which they may function as relative pronouns – next to their function as distal demonstratives – in accordance with the hypothesis in (299). Pronouns *deze* and *dit* only differ from *die* and *dat* in their specification for location as proximal, as a result of which they are always forced to have a deictic interpretation, and may not function as relative pronouns, i.e. they only function as proximal demonstratives.[19]

Table 4.4: Dutch *d*-pronouns

	pronoun			
	die	*deze*	*dat*	*dit*
number				
gender	common	common		
animacy				
location		proximal		proximal
referentiality	definite	definite	definite	definite

4.2.4 *w*-pronouns

As previously mentioned, a relative pronoun is *referential* or *anaphoric* in that it refers to "a concept that has been expressed elsewhere in the sentence" (Bhat 2004:230), namely the RC head. Whereas *d*-pronouns are inherently specified for referentiality, *w*-pronouns *wie* 'who' and *wat* 'what' are not. What needs to be explained is thus why *w*-pronouns can be used as referential relative pronouns. Bhat (2004:section 11.4) refers to this question – based on the observation that many languages use relative pronouns that have the same form as interrogative pronouns – as the *relative-interrogative puzzle*.

I believe the key to answering this question lies in the feature specification of *w*-pronouns. Recall that in chapter 2 I argued that *w*-pronouns *wie* and *wat* are underspecified for referentiality.[20] This is indicated in (319).

[18] Recall from chapter 2 that I follow a proposal by Kester (1996), Rooryck (2003), Toebosch (2007), van Kampen (2007) and Barbiers et al. (2009) amongst others, and take *neuter* to be the underspecification for gender in Dutch (cf. Harley and Ritter 2002 for a more universal claim).

[19] It is unclear at this point why generally only *die* and *deze* – but not *dat* and *dit* – can be used with plural antecedents/referents (but see section 2.5.2.5). More specifically, if the feature specifications in table 4.4 are correct, all *d*-pronouns are predicted to occur with plural antecedents/referents, as all *d*-pronouns are underspecified for number. I leave this issue for future research.

[20] In a more or less similar vein, van Riemsdijk (1978:39) notes that because *wh*-pronouns can be used both as interrogative and as relative pronouns, they are presumably unspecified for definiteness ([αDEF] in his terms).

(319) *w*-pronouns: [referentiality:]

When *w*-pronouns occur by themselves, they thus get an indefinite (or interrogative) interpretation. However, because they are underspecified for referentiality, *w*-pronouns can occur in environments like RCs as well. Put differently, *w*-pronouns are compatible with the referential meaning of a relative pronoun when they occur in a construction with a RC head. I leave the question of the exact role of referentiality (and definiteness) in the context of RCs to future research.

Finally, let me point out that Bhat (2004:266) makes the particularly interesting observation that "languages that use the same pronoun as relatives and interrogatives also show affinity between interrogatives and indefinites" (the Dutch pronoun *wat* 'what' is a case in point, cf. section 2.5.2.2). The latter observation is also known as the *interrogative-indefinite puzzle* (Bhat 2004:chapter10). Postma (1994) provides a solution to this puzzle by arguing that a *wh*-pronoun is an *open variable* (in the sense of Heim 1982) that receives its interpretation – interrogative or indefinite – configurationally. The interrogative reading arises through movement of the *wh*-pronoun to the left periphery, as a result of which it remains unbound and is dependent on its hearer for binding (which happens by means of an answer). The indefinite reading arises when the variable stays in-situ (within VP) and is existentially bound by *Existential Closure* at LF (Heim 1982). This solution to the interrogative-indefinite puzzle is not incompatible with anything I have said so far.

4.2.5 A cross-linguistic perspective

Van Kampen (2007) observes the following correlation (which is corroborated by the typological survey in de Vries 2002): the option of using A-bar *d*-pronouns as relative pronouns is restricted to V2 languages. Furthermore, of those V2 languages, only languages that exhibit grammatical gender distinctions may use relative *d*-pronouns. The West-Germanic languages may therefore be classified as in table 4.5.

Table 4.5: Relative pronoun selection in West-Germanic restrictive RCs

	syntactic gender	no syntactic gender
V2	*d*-pronouns e.g. Standard Dutch, Frisian, High German	*wh*-pronouns e.g. Afrikaans
no V2	**d*-pronouns[21]	*wh*-pronouns e.g. English[22]

Something additional needs to be said about the Scandinavian languages. They exhibit V2 and grammatical gender distinctions, but they generally do not have restrictive RCs introduced by *d*-pronouns (although *d*-relative pronouns are used to some extent, e.g. Norwegian and Danish *der*-relatives, cf. Smits 1988, Vikner 1991, Åfarli 1994 a.o.). Rather, the unmarked, most frequent relativization strategy in Scandinavian is one by which the invariant relativizer *som* (in Mainland Scandinavian) or *sem* (in Icelandic) introduces the RC. Interestingly, the Scandinavian languages all exhibit embedded V2 in the presence of a complementizer (cf. Vikner 1994 a.o.).[23] This property seems to be related to the absence of *d*-morphology on the complementizer – see section 4.3.1 for details – which in turn might be related to the absence of *d*-relative pronouns. Since at this point it is unclear how exactly the different parameters interact, I leave this issue to further research.

Let us go back to the languages that have *d*-relative pronouns. Provided that my claim that proximal demonstratives are specified for location holds universally, we make a very clear prediction: if a language has a relative pronoun that is taken from the demonstrative *d*-pronoun paradigm, this will never be the proximal demonstrative pronoun. This prediction is borne out for German, as illustrated in (320).[24]

[21]The fact that Romance languages make use of relative pronouns that are taken from the *wh*-pronoun paradigm might be explained by the observation that Romance languages lack V2 and therefore lack A-bar *d*-pronouns.

[22]For this to be the right classification of English, it needs to be assumed that *that* is a complementizer in English RCs (cf. Bresnan 1970 a.o., but contra e.g. Kayne 2008, 2010).

[23]Icelandic (and Yiddish) differs from Faroese and the Mainland Scandinavian languages in that embedded V2 plus complementizer occurs in all embedded clauses, whereas in Faroese, Norwegian, Swedish and Danish, embedded V2 plus complementizer only occurs under certain matrix verbs – just like embedded V2 without complementizer in German (Vikner 1994:130).

[24]Notice that German uses the definite determiner as relative pronoun, and not the distal demonstrative *jener*. Definite determiners in German can be used as pronouns (i.e. they can license an empty NP) because they show strong agreement (Wiltschko 1998). The German system is thus different from the Dutch system in that Dutch uses the distal demonstrative pronoun as relative pronoun whereas German uses the neutral definite determiner as relative pronoun. Even so, the generalization that elements that are specified for location (as *proximal*) cannot be used as relative pronouns is maintained. It might be the case that the distal demonstrative *jener* in German is specified for location as *distal*, as a result of which it cannot be used as relative pronoun.

(320) a. der Mann [$_{RC}$ <den/*diesen> ich gesehen habe]
the.M.NOM man.M that.M.ACC/this.M.ACC I seen have
'the man I have seen'

b. das Buch [$_{RC}$ <das/*dieses> ich gelesen habe]
the.N.NOM book.N that.N.ACC/this.N.ACC I read have
'the book I have read' [German]

Notice furthermore that if complementizers are derived from demonstrative/relative pronouns (cf. *infra*), and if *neuter* is the underspecified value for gender universally (cf. Harley and Ritter 2002), it follows from the underspecification hypothesis in (299) that the declarative complementizer in German is identical to the *neuter* (distal) determiner/pronoun *das*, just like Dutch *dat* is. I assume that the different orthography in German (*das* vs. *dass*) is misleading and does not reflect a difference between the two instances of *das*/*dass* (cf. Kayne 2011).

(321) ich glaube [$_{CP}$ <dass/*dieses> ich den Mann gesehen habe]
I believe that/this I the.M.ACC man.M seen have
'I believe I have seen the man' [German]

The same observation holds for English: the declarative complementizer or relative pronoun – for the moment I remain agnostic about the status of complementizer/pronoun *that* (but see section 4.4) – is identical to the *distal* demonstrative pronoun (not the *proximal* demonstrative pronoun, cf. Kayne 2010).

(322) a. the man <that/*this> I saw
b. I believe <that/*this> I saw the man.

4.3 On the nature of complementizers

It is well known that finite declarative complementizers often coincide with nominal elements: demonstratives in West-Germanic (e.g. Dutch *dat*, English *that*, German *dass*), and *wh*-phrases in Romance (e.g. Latin *quod*, French *que*, Spanish *que*, Italian *che*).[25] As mentioned by van der Horst (2008:279), whereas

[25]I do not have any insight to offer as to why the declarative complementizer in West-Germanic corresponds to a *d*-pronoun, whereas the declarative complementizer in Romance corresponds to a *wh*-pronoun, besides the aforementioned correlation between the presence of V2 and the presence of *d*-relative pronouns in a particular language. I leave this issue for further investigation. For the time being, suffice it to say that in both Germanic and Romance it is an underspecified pronoun (e.g. underspecified for gender) that is identical in form to the declarative complementizer (in line with the hypothesis in (299)).

Kayne (2011) takes complementizers to be demonstrative/relative pronouns (cf. section 4.4) and proposes to relate the difference between West-Germanic and Romance to the *obligatory* movement in Romance demonstrative DPs versus the absence of movement in Germanic demonstrative DPs. For some reason, movement inside the demonstrative DP – namely noun raising – is incompatible with further NP raising for relativization (cf. section 3.2.2 for the details of a head raising analysis of RCs). As Kayne notes himself, it is unclear why this should be the case (maybe it is due to *improper movement* or *freezing*), and as far as I can

it is commonly assumed that the Dutch complementizer *dat* is etymologically identical to pronoun *dat* (cf. Kiparsky 1995 a.o. for Germanic languages more generally), it is unclear how complementizer *dat* developed from pronoun *dat*. This has to do with the fact that complementizer *dat* (as well as pronoun *dat*) was already present from the beginning of the written tradition. Diachrony thus cannot offer any real insight into the relation between complementizer *dat* and pronoun *dat*. Therefore, in this section, I remain largely agnostic about the exact diachronic relation between complementizer *dat* and pronoun *dat* (but see Kiparsky 1995 and Roberts and Roussou 2003 a.o.). I only point out that if complementizer *dat* indeed developed from pronoun *dat* through a process of *grammaticalization* (from lexical to functional element, or, as in this case, from functional to higher functional element), it should not come as a surprise that it is the most underspecified *d*-pronoun that served as a source for grammaticalization (compatible with the hypothesis in (299)).[26] This idea receives support from the observation that exactly the same holds for English *that* and German *dass* (cf. section 4.2.5): it is the most underspecified pronoun (underspecified for location (and gender)) that is identical in form to the finite declarative complementizer.

So, although I think it is no accident that the finite declarative complementizer is identical in form to the most underspecified *d*-pronoun (cf. table 4.4), I will not go as far as claiming that they are the same lexical item. More specifically, I follow common tradition and take complementizer *dat* to be a (morphologically complex) head – in contrast to pronoun *dat*, which I argued in chapter 2 is the spell out of an XP.[27] I will furthermore argue at length that claiming that complementizer *dat* is the same lexical item as (relative) pronoun *dat* cannot be right (*pace* Kayne 2008, 2010 a.o.) as it runs into all sorts of theoretical and empirical problems.

4.3.1 Complementizers as (complex) heads

Complementizers of the *dat*-type are standardly taken to be heads (X^0s) that occupy a position in the extended projection of the verb (CP) in subordinate

see, the success of this proposal is dependent on the success of the raising analysis of RCs, or a head internal analysis of RCs more generally, which I argued against in chapter 3.

[26] I do not consider the correspondence between complementizer *dat* and pronoun *dat* to be the result of *accidental homophony*, as that hypothesis lacks explanatory power (cf. Manzini 2010 and see the arguments against a lexicalist approach to multipurpose pronouns in section 4.1).

[27] The claim that complementizer *dat* is a head (X^0) whereas pronoun *dat* spells out a phrase (XP), seems incompatible with the assumption that *dat* is listed in the lexicon only once (unless a dual X-bar status for a single lexical element is postulated). It might however be possible to account for the fact that a single lexical item may surface as a head *and* as a phrase in terms of the *Superset Principle* (cf. section 2.5.4 for details). According to the Superset Principle, a Vocabulary Item can spell out a syntactic structure that is smaller than itself. More specifically, *dat* could be listed as an XP in the lexicon, but in some cases this XP realizes a structure that is smaller than itself (a subconstituent), here an X^0. I will not pursue this line of reasoning here.

clauses – this is what I assumed for *dat* (and *of* 'whether') throughout chapter 3. One of the main reasons for this claim is that it immediately accounts for the complementary distribution of the declarative complementizer and the finite verb. That is, under the traditional view of V2 (cf. den Besten 1983, but contra Zwart 1993) both the complementizer and the finite verb target the C^0 position, hence the unavailability of embedded V2 with an additional complementizer in Dutch, as illustrated in (323).[28] Notice that this account of the complementary distribution of complementizer and finite verb only works if there is just one CP-projection.

(323) a. Ik geloof [$_{CP}$ **dat** hij het gedaan **heeft**].
I believe that he it done has
'I believe he has done it.' complementizer + V-final

b. % Ik geloof [$_{CP}$ hij **heeft** het gedaan].
I believe he has it done
'I believe he has done it.' embedded V2, no complementizer

c. * Ik geloof [$_{CP}$ **dat** hij **heeft** het gedaan].
I believe that he has it done
embedded V2 + complementizer

Leu (2010) makes a particularly interesting observation regarding the structure of complementizers in Germanic (cf. Postma 1997 for the same observation, but a slightly different implementation).[29] He notices a strong correlation between the form of the declarative complementizer in Germanic (with or without *d*-morpheme) and the presence or absence of embedded complementizer V2 (henceforth eCV2), i.e. embedded V2 with a co-occurring complementizer, as illustrated in (324) for Danish.

(324) Vi ved [at denne bog har Bo ikke lest].
we know that this book has Bo not read
'We know that Bo did not read this book.'
[Danish, Leu 2010:10, adapted from Vikner 1995:66]

[28]The SAND2 data (Barbiers et al. 2008a:46) show that a couple of Northern Dutch varieties (8 out of 89 locations) do not obey this generalization: they allow embedded V2 in combination with a complementizer, as illustrated in (i) – cf. Hoekstra and Marácz (1989:84). Notice that as the complementizer in these varieties contains the *d*-morpheme, these data form counterexamples to the generalization put forward by Leu (2010), cf. *infra*.

(i) Jan vindt dat je moet zulke dingen niet geloven.
Jan thinks that you must.FIN such things not believe
'Jan thinks that you should not believe such things.'

[29]Instead of assuming that the *d*-morpheme starts out in a C-position (cf. *infra*), Postma (1997:3) takes the *d*-morpheme to be "an overt sign of the deictic properties that are supposed part of tense" and thus to start out in I^0/T^0.

This correlation is illustrated in table 4.6, which is copied from Leu (2010:11).[30,31]

Table 4.6: eCV2 and complementizer shape across Germanic

language	eCV2	complementizer
Afrikaans	no	d-at
Dutch	no	d-at
Frisian	no	d-at
German	no	d-ass
Yiddish	yes	-az
Danish	yes	-at
Faroese	yes	-at
Icelandic	yes	-að
Norwegian	yes	-at
Swedish	yes	-att

Both Postma (1997) and Leu (2010) notice that the generalization in table 4.6 receives support from diachronic change in Yiddish. In earlier stages of Yiddish (West Yiddish), the language had a complementizer introduced by a *d*-morpheme (*dz*/*daz*) and did not allow eCV2. Contemporary Yiddish (East Yiddish), on the other hand, has a complementizer that does not feature the *d*-morpheme and it allows eCV2. This simultaneous change strongly indicates that the two phenomena really go hand in hand.

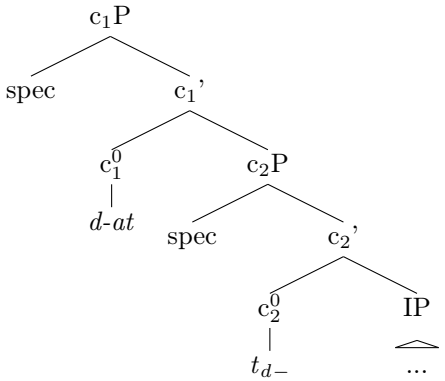
The *d*-morpheme and the finite verb in embedded clauses are thus in complementary distribution. Leu (2010) takes this to mean that either C^0 is lexicalized by *d-* (as in Afrikaans, Dutch, Frisian and German) or the finite verb moves to this C^0 position (as in Scandinavian and Yiddish). More specifically, because the verb in eCV2 constructions *follows* complementizer *at*, but the *d*-morpheme *precedes* complementizer *at*, Leu (2010) assumes a split CP domain and takes *d-* to be generated below *at*. In syntax, *d-* then moves to *at* (this assumption is in line with the mirror principle, cf. Baker 1985, Brody 2000). This contrast is illustrated in (325).[32]

[30]English is set aside by Leu (2010). Postma (1997) does consider English, and accounts for its behavior by adding another parameter: V-to-Agr movement. That is, whereas Dutch and English both have I/T-to-C movement (and a *d*-complementizer), English, in contrast to Dutch, does not have V-to-Agr movement (because it does not have a strong Agr).

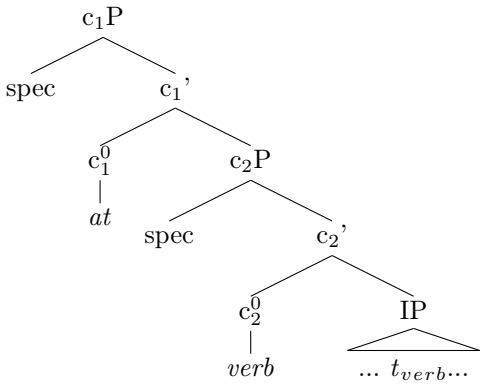
[31]The SAND1 data (Barbiers et al. 2005) show that some Dutch dialects have a declarative complementizer that does not feature the *d*-morpheme, e.g. *as*, *at*. Interestingly, this form of the complementizer only shows up in combination with an interrogative or relative pronoun that introduces the clause. Put differently, if a complementizer introduces the clause, it is invariably *da(t)* (there are only some exceptions in the Dutch speaking part of France, cf. SAND1 data (Barbiers et al. 2005:14)).

[32]In addition to the alternation involving eCV2 and the (non-)absence of the *d*-morpheme, Leu (2010) mentions two other alternations that relate to the presence or absence of this morpheme, namely strong adjectival agreement and definiteness marking, as illustrated in (i) and (ii) respectively. Leu (2010) assumes that the extended projections of V, N, and A, all have a complementizer head (*c*) in their left periphery, and that the three alternations

(325) a. lexicalization of c_2^0 by *d-* (Dutch, cf. Afrikaans, Frisian, German)



b. verb movement to c_2^0 (Danish, cf. Yiddish, Scandinavian)



Notice that if complementizer *dat* is indeed morphologically complex as indicated in (325a), the structure of the left periphery in Dutch is even more

involving the *d*-morpheme all reduce to one and the same contrast. Either *c* can be lexicalized by the *d*-morpheme (as in (ia) and (iia)), or there is (phrasal) movement of a lexical category into the *c*P domain (the adjective in (ib,c) and the noun in (iib)). Simply put: "*c* hosts *d-* or attracts *x*".

(i) a. **d-er** gut-e Wein
the-STR good-WK wine

b. ein gut-**er** Wein
a good-STR Wine

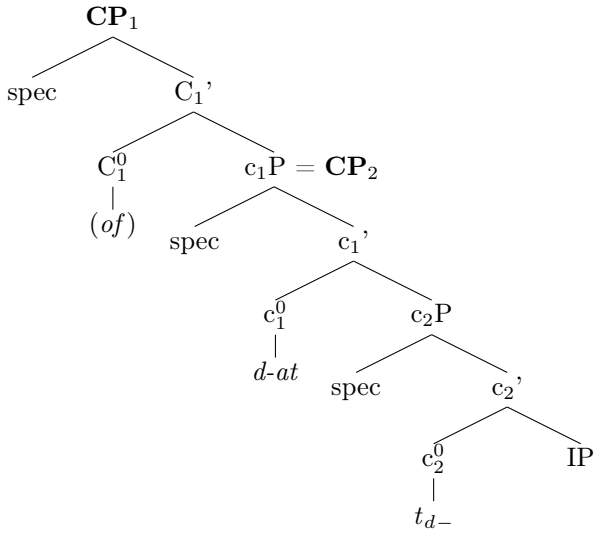
c. Ø gut-**er** Wein
good-STR Wine [German, Leu 2010:7-8]

(ii) a. d-**as** Haus
the-AGRN house [German, Leu 2010:12]

b. hus-**et**
house-AGRN [Danish, Leu 2010:12]

elaborate than assumed before.[33] More specifically, it looks something like the structure in (326). The highest CP layer is related to clause typing, similar to what I assumed in chapter 3. The lower CP layer, however, which I argued before is (optionally) headed by *dat*, should in fact be split further into a layer that hosts the *d*-morpheme and a layer that hosts the element *at*. The lower two layers of the CP domain *together* thus equal the CP_2 layer that I argued for in chapter 3.

(326) The structure of the left periphery in embedded clauses



4.3.2 The proposal

As for the feature specification of *dat*, recall from sections 2.5.2.5 and 4.2.3 that pronoun *dat* is only specified as *definite*, as given here in (327).

(327) feature specification of *dat*

a. [referentiality: definite] = definite
b. [number:] = underspecified for number
c. [gender:] = underspecified for gender
d. [animacy:] = underspecified for animacy
e. [location:] = underspecified for location

As mentioned earlier, in line with the hypothesis in (299), if the finite declarative complementizer is derived from a demonstrative/relative *d*-pronoun, it is

[33] The element *at* might even be further decomposed, e.g. into an invariant part *-a-* (potentially the *distal* morpheme, as argued by Rooryck 2003) and an inflectional part (cf. complementizer agreement).

expected that it is exactly the most underspecified *d*-pronoun that corresponds to this complementizer. Put differently, I believe it is the (almost complete) underspecification of *dat* that makes it suitable as a source for grammaticalization. If true, the natural question that arises at this point is whether it can be plausibly argued that complementizer *dat*, just like pronoun *dat*, is specified as definite as well – assuming that the *d*-morpheme expresses some sort of definiteness (cf. 2.5.2.1). This question becomes even more pertinent if we take into account the fact that *wat* is even more underspecified than *dat* – *wat* being underspecified for referentiality/definiteness – yet *wat* does not function as a complementizer. Put differently, what is the role of the *d*-morpheme in the Dutch finite declarative complementizer?

I follow Roberts and Roussou (2003) (see also Postma 1997, Hoekstra 2004, Leu 2008 a.o.) and take the (*d*-morpheme on) the complementizer to indeed express some sort of *definiteness* or *deixis* (the [+declarative] feature in their terms):

> *[it] can be taken as deictically referring to the truth of the proposition expressed by the IP complement to C [...] in the same way that the demonstrative deictically refers to the individual expressed by the complement to D* (Roberts and Roussou 2003:114)

Roberts and Roussou furthermore claim that the difference in meaning between the demonstrative pronoun and the complementizer comes from the different complements they take. When it takes an NP as its complement, it denotes an *entity*, and when it takes a TP as its complement, it denotes a *proposition*.[34,35] As a complementizer, *dat* thus introduces a propositional complement just like pronoun *dat* introduces a noun phrase. This similarity between DP and CP fits in nicely with the widely recognized (structural) parallelisms between DP and the clause (Abney 1987, Szabolcsi 1987, 1994, Cardinaletti and Starke 1999 among many others).

[34]Notice that in my proposal, the demonstrative determiner *dat* strictly speaking does not take an NP as its complement, as it is a phrase in SpecDP (cf. footnote 13 of this chapter).

[35]Similarly, Manzini and Savoia (2003) and Manzini (2010) (and cf. Roussou 2010) take complementizers to *be* nominal, the difference between a pronoun and a complementizer being whether it ranges over individuals or propositions. More specifically, Manzini (2010) argues that declarative complementizers of the Italian *che*-type are not functional categories of the verb but independent nominal heads that are selected by the higher (matrix) verb and that take the embedded clause as their complement. As a result of this, the C position in the embedded clause is exclusively verbal and can only be targeted by a verb. As mentioned by Manzini (2010:180) herself, theories that assume a *split CP* structure (cf. Rizzi 1997, and see chapter 3) or *CP recursion* have the same empirical coverage. For example, both theories predict the existence of embedded V2 under a complementizer. However, such theories need to say something additional about the homophony between complementizers and pronouns.

Assuming that sentential complementation, just like regular DP arguments, has some sort of nominal structure raises all sorts of questions, e.g. questions regarding selection and (islands for) extraction. Most of these questions will be addressed in section 4.4, when I discuss a proposal by Kayne (2008, 2010) that shares some of the insights of Manzini's proposal.

In sum, although I do not take complementizer *dat* and pronoun *dat* to be the same lexical item – they have a different X-bar status – I do take complementizer *dat* to be diachronically related to pronoun *dat*. The idea that the former is derived from the latter, I proposed, is the result of the highly underspecified nature of pronoun *dat*. If this line of reasoning about complementizer *dat*, as well as the feature specification of *d*-pronouns as given in table 4.4, is on the right track, we might expect that the next most underspecified *d*-pronoun, namely *die*, can act as a complementizer in certain environments as well. In section 4.5, I will indeed show in detail that in some varieties of Dutch in (long-distance) RCs with a common gender antecedent, *die* can be an (agreeing) complementizer as well, thereby strengthening the proposal made here.

So far, I have not said much about why complementizer *dat* should not be equated with (relative) pronoun *dat*, except for their different X-bar status. A recently popular view among linguists, notably Kayne (2008, 2010) and Arsenijević (2009), is that *all* complement clauses are in fact (a kind of) RCs,[36] as a result of which the difference between complementizers and relative pronouns becomes less clear.[37,38] Kayne (2008, 2010) even explicitly argues that the sentential complementizer is in fact a relative pronoun. The next section briefly reviews such proposals and shows that they run into all sorts of theoretical and empirical problems.

4.4 On complementation as relativization

In his 2008 paper, Kayne claims that as a consequence of Antisymmetry (in the sense of Kayne 1994) there exist two disjoint types of lexical items in the language faculty: *nouns* and *non-nouns*. Kayne argues that only *non-nouns* can project, as a consequence of which nouns can exclusively *be* complements but not *have* them. This claim immediately raises the question about the nature of *that they're here* in constructions like *the fact that they're here* in (328a), which under traditional assumptions is simply a complement to the noun *fact*. Kayne claims that such structures involve a RC structure. The fact that (328a) differs from familiar RCs as in (329) – in the sense that *which* is not possible in the former as can be seen in (328b) – Kayne explains by arguing that (328a) is the result of relativizing the object (*fact*) of a (silent) preposition *in*. More specifically, the structure in (328a) is the result of relativizing *fact* in constructions like (330a), as illustrated in (330b); silent elements are indicated by small caps.

[36]This idea has its roots in the work of Rosenbaum (1967).

[37]There is a long tradition of linguists taking *factive complements* to be (a kind of) RCs (i.e. clauses with a nominal head), starting with Kiparsky and Kiparsky (1971).

[38]See Haegeman and Ürögdi (2010) amongst others for the claim that although relativization is pervasive in (certain types of) sentential complementation, relativization is not to be equated with complementation.

(328) a. the fact *that* they're here
b. * the fact *which* they're here

(329) a. the fact *that* they mentioned
b. the fact *which* they mentioned

(330) a. they're here in fact/ they're in fact here/ in fact they're here
b. the fact$_1$ *that* they're here IN _$_1$

That RCs may contain a silent preposition is illustrated by (331a), which is a RC structure that arguably contains a silent *in* by analogy with (331b) – the lack of a visible preposition is thus not specific to RCs.[39] Crucially, whenever there is a silent *in*, *which* becomes impossible, as illustrated in (331c). So, the observation that *which* is not possible in (328b) is a result of the fact that silent *in* is incompatible with *which* independently. Put differently, taking the sentences in (328) to contain a silent preposition *in*, (328b) is ungrammatical for the same reasons (331c) is: *which* and silent preposition *in* are mutually exclusive. Although Kayne relates the ungrammaticality of (331d) to the extra restrictions that are posed on determiners found with *fact* – e.g. (332a) and (333a) vs. (332b) and (333b) respectively – it remains unclear why *exactly* this sentence is ungrammatical.[40]

(331) a. the way (that) they solved it
b. they solved it (in) this way
c. * the way which they solved it
d. * the way in which they're here

(332) a. In <what/?which> way did they solve it this time?
b. * In <what/which> fact are they here this time?

(333) a. We solved it in another way.
b. * We're here, in another fact.

[39]The fact that preposition *in* can be spelled out in (331b) but not in (331a) – **the way in that they solved it* – Kayne explains by the following descriptive generalization, that is based on a cross-linguistic comparison between Dutch, German and English.

(i) In (at least) West-Germanic, a demonstrative-related relative pronoun can be the object of an adposition only if that *d*-pronoun has morphological Case
[=(35) in Kayne 2010:196]

[40]I leave aside here the discussion on *derived nominals* and *possessive* structures, i.e. the status of *evidence* in constructions like *the removal of the evidence* and the status of *yours* in constructions like *a friend of yours* respectively – they cannot be complements according to Kayne's (2008) proposal – as it is irrelevant to the discussion on sentential complementation as relativization.

Building in part on Kayne (2008), Kayne (2010) claims that sentential complementation involves (partly invisible) RCs as well. He furthermore argues that what has always been taken to be a complementizer, namely *that* that introduces sentential complements, is in fact a relative pronoun (cf. Manzini 2010 a.o. who also argues against elements like *that* to be complementizers in the traditional sense, see section 4.3.1).[41]

Kayne (2010) starts by claiming that analyses like Roberts and Roussou (2003) which take complementizer *that* to have originated diachronically from demonstrative *that* (cf. section 4.3.1), wrongly ignore the important observation that *this* did not provide a source for a complementizer – (334a)-(334b) vs. (334c)-(334d) – as well as the question of why this should be so.[42]

(334) a. We say *that*: the earth is round.
b. We say *that* the earth is round.
c. We say *this*: the earth is round.
d. * We say *this* the earth is round.

[41] In this footnote, I briefly mention four arguments against *that* to be a relative pronoun – originally proposed by Klima (1964) to account for the different behavior of *that* in relation to relative pronouns *who* and *which* – as well as the refutation of these arguments as proposed by Kayne (2010:193-202).
(i) Relative *who*, unlike relative *that*, can pied-pipe the preposition of which it is the object (*the person to <whom/*that> we were alluding*). Kayne (2010): *that* cannot be the object of a preposition for independent reasons: it is a demonstrative-related relative pronoun without morphological case (cf. the descriptive generalization in footnote 39).
(ii) Relative *who*, unlike relative *that*, can pied-pipe a larger DP of which it is the possessor (*the person <whose/*that's> book we were talking about*). Kayne (2010:194): this reduces to the fact that "demonstratives accompanied by no overt NP cannot be possessors" (**that's importance is undeniable*).
(iii) Relative *that* shows no number agreement (despite a plural form *those* that is present elsewhere in the grammar and despite the existence of a plural form of the relative pronoun in related languages). Kayne (2010:201): this reduces to the fact that "English plural *-s* can be prenominal only if it is also pre-D": *those D books* vs. *D interesting(*s) books*. A sentence like **the only books*$_1$ *those D* t_1 *you should read* with pre-D *those* is ungrammatical because the relative head *books* cannot raise to SpecDP to establish a RC structure, because the demonstrative *those* already occupies SpecDP (English non-agreeing relative *that* thus most likely occurs in post-D position).
(iv) Relative *that*, unlike relative *who* and *which*, appears to be indifferent to the +/-human nature of its antecedent (*the person <who/*which/that>, the house <*who/which/that>*). Kayne (2010): this claim is misleading, because *that* is in fact sensitive to the +/-human nature of its antecedent, e.g. in a cleft sentence like *in fact it was Mary, <who/?*that> got me interested in linguistics in the first place*.

[42] Although it might be possible to argue for an SVO language like English that the grammaticalization from demonstrative *that* (334a) to complementizer *that* (334b) relates to the change from *parataxis* to *hypotaxis* (cf. Kiparsky 1995), it is less clear that the same can be argued for an SOV language like Dutch, because the word order in matrix clauses (V2) differs from the word order in embedded clauses (V-final). This is illustrated in (i).

(i) a. We zeggen *dat*: de aarde is rond.
we say that the earth is round
b. We zeggen *dat* de aarde rond is.
we say that the earth round is

Kayne argues that there is no such thing as a complementizer in the standard sense of the term, and that sentential *that* is an instance of demonstrative/relative pronoun *that*. Explaining the absence of complementizer *this* then reduces to explaining the absence of relative pronoun *this*. As *this* seems to be necessarily deictic (context-sensitive), whereas *that* can be more neutral, Kayne assumes that *this* is necessarily accompanied by a silent first person element (akin to *I/me*) in its specifier. This person element blocks raising of the relative head noun to SpecDP, as a consequence of which the structure will never converge – multiple specifiers are disallowed in Kayne's (1994) Antisymmetry framework. This is illustrated in (335).[43]

(335) a. the only [books$_1$ *that* _$_1$] you should read
b. * the only [[$_{SpecDP}$ books$_1$ PERSON] *this* _$_1$] you should read

Assuming then that both sentences in (336) are RCs, immediately accounts for their ungrammaticality: *this* cannot be a relative pronoun.

(336) a. * the fact *this* they're here
b. * the fact *this* they mentioned

Similarly, sentential complements of factive predicates are taken to be RC structures as well, thereby immediately explaining the ungrammaticality of *this* in such structures, cf. (337a). More specifically, inspired by Kiparsky and Kiparsky (1971), factives are assumed to involve a silent or deleted FACT that acts as the RC head. Factives are thus RCs based on silent IN FACT, as illustrated in (337b). As mentioned by Kayne (2010), that this line of reasoning is on the right track is suggested by the fact that in Modern Greek the RC marker *pou/pu* is also used to introduce factives (Roberts and Roussou 2003, Roussou 2010, and cf. Krapova 2010 for Bulgarian *deto*).

(337) a. We're sorry <that/*this> you're here.
b. We're sorry FACT$_1$ [that you're here IN _$_1$].

Following an insight by Rosenbaum (1967) who proposes that sentential complements and sentential subjects are all accompanied by (a silent) *it*, Kayne argues that *all* sentential complements – not only factives – involve a head noun. As for non-factive sentential complements, either (silent) *it* or a silent noun will be raised from inside the RC to a left peripheral position.

Although Kayne's (2008, 2010) proposal has the great advantage of being able to immediately account for the homophony between complementizers and relative pronouns, and for the absence of complementizer/relative pronoun *this*, the proposal has not (yet) been worked out very well. It sidesteps several important issues, thereby raising many questions and leaving many unresolved

[43] Notice that this account for the lack of *this* as a relative pronoun (and consequently as a complementizer) is in large part dependent on the success of the *raising* analysis of RCs as in Kayne (1994) – which I argued against in chapter 3.

puzzles. First and most obviously, by taking (sentential) complementation to involve (*raising*) RC structures, all types of embeddings have the same underlying structure and are predicted to behave exactly the same, as a consequence of which a straightforward explanation of all sorts of asymmetries between complements and adjuncts/RCs is lost.[44] It is for example well known that complements and adjuncts/RCs to nouns behave differently with respect to reconstruction, as illustrated here for Condition C. Notice that it is irrelevant to this observation how connectivity effects are best accounted for, but cf. the discussion in chapter 3.

(338) a. * Which claim [that John$_i$ was asleep] was he$_i$ willing to discuss?
b. Which claim [that John$_i$ made] was he$_i$ willing to discuss?

Furthermore, as noted by Koster (2000a) amongst others for Dutch, the extraposition behavior of sentential complements is different from that of RCs: whereas sentential complements generally *obligatorily* extrapose, as illustrated in (339), extraposition from NP is *optional*, as illustrated in (340).[45] This brings about the conclusion that "it seems unlikely that we have to do with a unitary phenomenon" (Koster 2000a:3).

(339) a. Jan heeft beweerd [$_{CP}$ *dat* zij zou komen].
Jan has claimed that she would come
'Jan claimed that she would come.'
b. * Jan heeft [$_{CP}$ *dat* zij zou komen] beweerd.

[44]For present purposes it is irrelevant how exactly the difference between arguments and adjuncts/RCs is modeled in the model of grammar. Even though in my proposal RCs are generated as complements (to D^0), and not as adjuncts (cf. chapter 3), the important thing to notice is that their syntax is crucially different from the syntax of complements, whereas in Kayne's proposal the syntax of complementation is identical to the syntax of relativization.

[45]This is a bit of an oversimplification, as can be seen in (i) and (ii). As shown by Barbiers (2000), whereas non-factive/propositional CP complements obligatorily extrapose (i), factive CP complements may also occupy a position in the middlefield (ii); the acceptability of (iib) is not completely uncontroversial, hence the % in front of it. Factive CP complements thus seem to behave more like RCs (and DPs more generally) than non-factive/propositional CP complements do (cf. Kiparsky and Kiparsky 1971 and many others since). The examples in (i)-(ii) are taken from Barbiers (2000:192).

(i) a. Jan zal wel vinden [$_{CP}$ dat Piet geschikt is].
Jan will certainly find that Piet eligible is
'Jan will have the opinion that Piet is eligible.'
b. * Jan zal [$_{CP}$ dat Piet geschikt is] wel vinden.
Jan will that Piet eligible is certainly find

(ii) a. Jan zal nooit toegeven [$_{CP}$ dat ie gelogen heeft].
Jan will never admit that he lied has
'Jan has lied and he will never admit that.'
b. % Jan zal [$_{CP}$ dat ie gelogen heeft] nooit toegeven.
Jan will that he lied has never admit
'Jan has lied and he will never admit that.'

(340) a. Zij heeft de bewering [$_{CP}$ *die* Jan deed] gehoord.
she has the claim RP Jan did heard
'She heard the claim that Jan made.'

b. Zij heeft de bewering gehoord [$_{CP}$ *die* Jan deed].

Another important asymmetry that is not accounted for by Kayne's analysis is that between factive and non-factive predicates. Semantically, factive predicates, unlike non-factive predicates, *presuppose* the truth of their complements (Kiparsky and Kiparsky 1971). It is well known that factive predicates behave differently syntactically from non-factive predicates in several ways as well (cf. footnote 45), as a consequence of which it is generally assumed that factive and non-factive predicates differ *syntactically* (cf. Barbiers 2000 for Dutch: factive CPs are adjuncts and non-factive/propositional CPs are complements). For example, only factive predicates can have the noun *fact* as their complement (341), and only factive predicates can freely take gerunds as their objects (342); both examples are taken from Kiparsky and Kiparsky (1971:347).

(341) a. I want to make clear the fact that I don't intend to participate.
You have to keep in mind the fact of his having proposed several alternatives. [factives]

b. * I assert the fact that I don't intend to participate.
* We may conclude the fact of his having proposed several alternatives. [non-factives]

(342) a. Everyone ignored Joan's being completely drunk.
I regret having agreed to the proposal. [factives]

b. * Everyone supposed Joan's being completely drunk.
* I believe having agreed to the proposal. [non-factives]

Moreover, pronominalization of the sentential complement as *it* is grammatical in case of factive predicates, as illustrated for English in (343a), but ungrammatical in case of non-factive predicates – "although the difference is a delicate one, and not always clearcut" (Kiparsky and Kiparsky 1971:361) – as illustrated in (343b).[46] Although these facts seem to suggest that an analysis of sentential complements of *factive* predicates in terms of relativization is in fact not so far fetched (cf. Kiparsky and Kiparsky 1971 and many others since, but see e.g. Haegeman and Ürögdi 2010 for arguments against the claim that factive complements are *nominal*), these data make the claim that a (silent) *it* is relativized in non-factive sentential complements not very believable.

[46] According to Haegeman and Ürögdi (2010) the presence of *it* is not a proper diagnostic for factivity (nor for *contextual givenness*). This is illustrated by (i), in which the presence of *it* does not necessarily signal factivity. Rather, Haegeman and Ürögdi assume that the presence of *it* is related to *D-linking*.

(i) John was the most horrible boyfriend, who told me one lie after another. Yet, whenever he told me, I believed (**it**) that he would marry me. What an idiot I was! [Haegeman and Ürögdi 2010:143]

(343) a. Bill resents *it* that people are always comparing him with Mozart.
They didn't mind *it* that a crowd was beginning to gather in the street. [factives]

b. * Bill claims *it* that people are always comparing him to Mozart.
* They supposed *it* that a crowd was beginning to gather in the street. [non-factives, Kiparsky and Kiparsky 1971:361]

Furthermore, taking sentential complements (and complements to nouns) to be RCs makes them into Complex NPs, as a consequence of which we expect all sentential complements to be *strong islands* for extraction, contrary to fact. Put differently, we expect the sentences in (344) to have the same status: both involve a RC out of which a *wh*-pronoun extracts, in violation of the CNPC (Ross 1967).

(344) a. * Who$_1$ did you hear [$_{NP}$ the rumor [$_{CP}$ that Mary kissed $\__1$]]?
b. Who $_1$did you hear [$_{NP}$ IT [$_{CP}$ (that) Mary kissed $\__1$]]?

More generally, the embedded clause of *all* long-distance A-bar dependencies (introduced by *that*) is always predicted to be an island for extraction. Consequently, Kayne is forced to say that extraction out of RCs is allowed, thereby losing an account of the CNPC effects.

Explaining complement/adjunct asymmetries, factive/non-factive asymmetries, and CNPC effects in other than structural terms is not inconceivable – one could for example assign special properties to the different (silent) heads (e.g. overt RC head, silent IT, silent FACT) – yet weakens the argument. The burden of proof lies with theories like that of Kayne (2008, 2010) to show that the same asymmetries and CNPC effects can be mimicked by a theory according to which *all* (sentential) complements involve RCs.

Second, following Kayne's suggestion that "the claim that English sentential *that* is a relative pronoun must be taken to extend to (non-prepositional) finite complementizers in other languages" (Kayne 2010:19) by extending his analysis to Dutch, immediately raises the issue of doubly filled COMP (DFC) phenomena (in RCs; cf. chapter 3). Under traditional analyses DFC involves the simultaneous spell out of a pronoun and a complementizer in the left periphery (COMP) of a clause, but under Kayne's analysis DFC in RCs must involve the presence of two relative pronouns. As argued by Kayne (2010) towards the end of his paper, DFC is only an apparent problem to his theory. The argumentation for this claim is based on French DFC structures, as in (345).

(345) la fille **à qui que** tu as parlé
the girl to who what you have spoken
'the girl to whom you have spoken' [French]

The structure in (345) is claimed to be a *reduced* relativized cleft. The examples in (346a) and (346b) illustrate a relativized cleft and a non-relativized cleft in English respectively. The element *that* in both examples in (346) is the same *that*. Similarly, *que* in (347) is the same *que* as in (345).

(346) a. ? the person **to whom** it was **that** we were alluding
b. it was to him **that** we were alluding

(347) c'est à elle **que** tu as parlé
it is to her that you have spoken [French]

Whereas this reduced cleft analysis more or less seems to work for similar sentences in Dutch, as illustrated in (348),[47] DFC in RCs as in (349) and DFC in embedded *wh*-Qs as in (350) cannot be accounted for by this analysis, as such sentences do not (always) give the same DFC pattern *with* and *without* the *het is/was* 'it is/was' part.[48] The judgments on the sentences in (348)-(350) are from the MPQ2-B data.[49]

(348) a. % de man **met wie** het was **dat** zij gepraat hebben
the man with who it was that they talked have
b. % de man **met wie** ~~het was~~ **dat** zij gepraat hebben
the man with who that they talked have

(349) a. de man **die** het was <?***dat**/%**die**> wij geroepen hebben
the man *die* it was that/*die* we called have
b. de man **die** ~~het was~~ <?***dat**/***die**> wij geroepen hebben
the man *die* that/*die* we called have

[47] The sentence in (348a) is attested considerably less frequently than the sentence in (348b): 16% vs. 45% respectively (MPQ2-B data).

[48] The observation that sentences like *de man **die die** zij geroepen hebben* (cf. (349b)) do not occur, might get an independent explanation in terms of *haplology* – the tendency to avoid (accidental) repetition of identical morphemes within a particular syntactic environment (cf. Neeleman and van de Koot 2006) – as a result of which only one of the two instances of *die* is maintained. However, I would not be surprised to find sentences like *de man **wie** het is **die** zij geroepen hebben* and *de man **die** het is **wie** zij geroepen hebben*, in which case the pronouns are not spelled out identically and are thus not subject to haplology. If such sentences lie at the basis of DFC patterns, we predict DFC patterns like ***wie die*** and ***die wie***, but DFC patterns like that are not attested so far (they have not been explicitly tested in the MPQs). This might provide another argument against Kayne's analysis of DFC patterns, but further empirical research is required to settle this issue.

[49] Recall that the MPQ data underrepresent the Dutch speaking part of Belgium, i.e. the geographic spread of the informants is almost entirely restricted to the Netherlands (cf. section 1.3.2). For some of the sentences in (349) and (350), the SAND1 data therefore show a slightly different pattern. First, whereas the MPQ2-B data show that the sentence in (349a) with the combination *die dat* occurs only very marginally (28/380=7%), the SAND1 data (Barbiers et al. 2005:84) show that DFC in a RC with a *d*-pronoun is in fact attested in a northern area and a southern area. More specifically, the form *die-t* is attested in most locations in Friesland, and the combination *die as/at/da* is primarily attested in the northeast of Oost-Vlaanderen and the western part of the province of Antwerpen. Second, whereas the MPQ2-B data do not show a clear geographic distribution of the sentence in (350b) with the string *wie dat*, the SAND1 data (Barbiers et al. 2005:16) show that this DFC pattern is primarily attested in the Dutch speaking part of Belgium. More specifically, the presence of the declarative complementizer following the *wh*-pronoun is often obligatory in this part of the language area (cf. also Boef to appear).

(350) a. ik vroeg haar **wie** het is <?***dat**/%**die**> zij kent
I asked her who it is that/*die* she knows

b. ik vroeg haar **wie** ~~het is~~ <%**dat**/??**die**> zij kent
I asked her who that/*die* she knows

Third, it is not completely clear what governs the distribution of pronouns, most prominently *optional* versus *obligatory* pronouns. It is for example well known that complements to nouns do not easily allow *that*-deletion, whereas RCs and sentential complementation do (there furthermore seems to be a difference in *that*-deletion between factive and non-factive complements as well, in that the latter are more tolerant towards *that*-deletion, cf. Hoekstra 2004:193 a.o.). This is illustrated in (351).

(351) a. John saw the cup (that) Mary bought yesterday.
b. John said (that) Mary was ill.
c. The fact ??(that) Mary was ill surprised them all.
[Arsenijević 2009:43]

The asymmetric behavior between RCs and sentential complements with respect to *that*-deletion is further illustrated in (352) (the example in (352b) illustrates the *that-t* effect).

(352) a. the man *(that) kicked me
b. who do you think (*that) kicked me?

The patterns in (351) and (352) are mysterious under Kayne's analysis of these sentences, as they all supposedly involve relativization and should thus show the same grammaticality judgments regarding *that*-deletion.

In another attempt to equate (sentential) complementation with a type of RC, Arsenijević (2009) – without explicitly attempting to do so – solves some of the problems encountered by Kayne (2008, 2010). Arsenijević argues that all complement clauses involve relativization, the difference between finite complement clauses (FCC) and traditional RCs being that in the latter a nominal expression is relativized, whereas in FCCs *ForceP* is the relativization site. Put differently, the relativization site of FCCs is in the projection in the left periphery that specifies illocutionary Force.[50] More specifically, the subject of Force is assigned a variable value – [Var], cf. the semantic analysis of relativization as proposed in Adger and Ramchand (2005) – which is matched by a specified force value on the nominal head. The resulting structural relation between the FCC and the nominal head is that of relativization. This is illustrated for a FCC to a nominal expression in (353).

[50]Notice that it is not typically assumed that embedded clauses have illocutionary force (cf. Zanuttini and Portner 2003, Bayer 2004 amongst others).

(353) a. the claim that Mary came late

b. [$_{DP}$ the claim$_{[force:claim]}$ [$_{ForceP}$ [$_{SpecForceP}$ [Var]] that [$_{IP}$ Mary came late]]]

In case of FCCs to verbal expressions, Arsenijević assumes that an incorporated light nominal mediates the relation between the verb and the FCC, as illustrated in (354).

(354) a. John claimed that Mary came late.
= John [$_{VP}$ made [$_{DP}$ claim [$_{FCC}$ that Mary came late]]]

b. Mary believed that John was the murderer.
= Mary [$_{VP}$ held/had [$_{DP}$ belief [$_{FCC}$ that John was the murderer]]]

c. John asked whether Mary was the murderer.
= John [$_{VP}$ asked [$_{DP}$ question [$_{FCC}$ whether Mary was the murderer]]]
[Arsenijević 2009:43]

Differences between RCs and FCCs (e.g. transparency to syntactic operations, cf. *supra*) can be reduced to their different structural properties, e.g. what has been relativized (nominal expression or Force), and/or the behavior of the nominal (incorporation into the verb or not). This option is not immediately available under Kayne's approach to the identification of complementation and relativization, as RCs and FCCs are taken to be structurally identical. However, as far as I can see, it is unclear how differences within the realm of sentential complements, e.g. factive vs. non-factive predicates, are to be accounted for by Arsenijević's approach.

Haegeman and Ürögdi (2010) account for this difference by arguing that only *referential CPs* (RCPs) – roughly *factive* clauses[51] – are derived by relativization of an event operator (comparable to Arsenijević's proposal).[52] More specifically, they argue that RCPs involve operator movement – illustrated in (355) – as a consequence of which they are weak islands for extraction and resist *Main Clause Phenomena* (MCP) like argument fronting. *Non-referential* CPs (NCPs) – roughly *non-factive* clauses – on the other hand, do not involve operator movement, are not islands for extraction and are compatible with MCP. The different properties of RCPs and NCPs are illustrated in (356) and (357) respectively, in which the a-examples illustrate the islandhood of the CPs, and the b-examples show the possibility of argument fronting (a MCP).

(355) structure of referential CPs (RCPs): event relativization
[$_{CP}$ Op_1 C [$_{FP}$ t_{Op} [$_{TP}$ ...]]]

[51] As argued by de Cuba and Ürögdi (2010), *factivity* is a *lexico-semantic* rather than a *syntactic* property. *Referentiality* is in fact active in syntax, and it is exactly the referential nature of a complement clause that is responsible for its syntactic properties.

[52] Cf. movement derivations for temporal and conditional adverbial clauses (Bhatt and Pancheva 2006, Haegeman 2009, 2010 a.o.).

(356) *referential CPs (RCPs)*
  a. * How did you notice [that Maria fixed the car]?
  b. * John regrets that *this book* Mary read.

(357) *non-referential CPs (NCPs)*
  a. How do you suppose [that Maria fixed the car]?
  b. John thinks that *this book* Mary read.

Haegeman and Ürögdi do not take RCPs to be nominal – *pace* Kiparsky and Kiparsky (1971) a.o. – because RCPs do not consistently pattern with DPs. For example, in contrast to RCPs, DPs can be the complement of a preposition, as illustrated in (358).[53] Furthermore, in contrast to DPs (359c,d), RCPs can be the complement of adjectives (359a) and RCPs can be extraposed and associated with *it* (359b). The data in (358) and (359) are from Haegeman and Ürögdi (2010:136), the labels are mine. Notice that these data show another asymmetry that needs to be addressed and accounted for by analyses like the ones by Kayne and Arsenijević.

(358) a. John forgot/thinks [$_{RCP}$ that Jane left too early].
  b. John forgot/thinks [$_{PP}$ about [$_{DP}$ Jane's early departure]].
  c. * John forgot/thinks [$_{PP}$ about [$_{RCP}$ that Jane left too early]].

(359) a. I was surprised [$_{RCP}$ that he left].
  b. *It* was surprising [$_{RCP}$ that he left].
  c. * I was surprised [$_{DP}$ his departure].
  d. * *It* is surprising [$_{DP}$ his departure].

As neiter Arsenijević (2009) nor Haegeman and Ürögdi (2010) assume that complementizers need to be equated with relative pronouns – Arsenijević explicitly retains the traditional distinction between *complementizers* and *pronouns* – I will not further discuss these proposals here, because they are immaterial to the nature of complementizers and their relation to pronouns.

In sum, I showed that equating (all) complementation with relativization, and more importantly, equating complementizers with relative pronouns, leads to theoretical as well as empirical problems. In the next section, I will illustrate this further by providing a case study that shows that identifying complementizers as relative pronouns cannot be correct. This case study focuses on microvariation in Dutch long-distance RCs, with special attention on 'special *die*'.

[53] See Haslinger (2007) for cases like (358c) in Dutch.

4.5 Case study: special *die*

This section explores microvariation in Dutch long-distance RCs – with primary focus on the nature of element *die* – as a case study to show that complementizers cannot be equated with relative pronouns. Put differently, I provide an elaborate argument for the claim that complementizers should be distinguished from relative pronouns (*pace* Kayne 2008, 2010 a.o.). The fact that some elements may act as complementizers *and* as pronouns should be accounted for in terms of *grammaticalization*, i.e. reanalysis from functional elements into higher functional elements (cf. *supra*).

I will focus in detail on the nature of the element *die*, and I will show that in addition to being a multipurpose pronoun (cf. *supra*), in specific environments in certain varieties of Dutch, *die* can act as a complementizer as well. This is not unexpected, given its feature specification and position in the *d*-pronoun paradigm, cf. tables 2.2 and 4.4 which are repeated here as tables 4.7 and 4.8 respectively.

Table 4.7: Functions of Dutch A-bar pronouns *die*, *dat*, *wie* and *wat*

	d-pronouns		*w*-pronouns	
	die 'that.C'	*dat* 'that.N'	*wie* 'who'	*wat* 'what'
determiner/nominal modifier	+	+	−	+
demonstrative pronoun	+	+	−	−
relative pronoun	+	+	+	+
interrogative pronoun	−	−	+	+
exclamative marker	−	−	−	+
indefinite pronoun	−	−	−	+
resumptive pronoun	+	+	−	−
complementizer	+*	+	−	−

*only in certain southern Dutch varieties (this section)

Table 4.8: Dutch *d*-pronouns

	pronoun			
	die	*deze*	*dat*	*dit*
number				
gender	common	common		
animacy				
location		proximal		proximal
referentiality	definite	definite	definite	definite

As already mentioned in section 2.3.1, some Dutch dialects show subject/object asymmetries in forming long-distance relativization structures. Specifically, whereas in (long-distance) object relativization the most deeply embedded

clause is introduced by the declarative complementizer *dat* 'that', in the most deeply embedded clause of (long-distance) subject relativization, the element *die* shows up – 'special *die*', cf. (57), here repeated as (360) – as indicated abstractly in (361).

(360) **special *die***: *die* in the left periphery of the most deeply embedded clause in (long) subject RCs in dialects that show a s/o asymmetry

(361) a. D^0 [$_{CP}$ RC head Op_{REL}/rel.pronoun ... [$_{CP}$ **die** ...]]
subject extraction

b. D^0 [$_{CP}$ RC head Op_{REL}/rel.pronoun ... [$_{CP}$ **dat** ...]]
object extraction

This section proposes an analysis for this subject/object asymmetry (henceforth s/o asymmetry) in terms of micro-parameters, related to *complementizer agreement*.[54] More specifically, I will argue that special *die* is not a (weak) relative pronoun, as argued for in Bennis and Haegeman (1984), and recently in Koopman and Sportiche (2008), and Sportiche (2011).[55] Rather, I take special *die* to be a form variant of the declarative complementizer *dat* (in line with traditional analyses of the French *que/qui* alternation, cf. Kayne 1976, Rizzi 1990 a.o.).

4.5.1 The data

In Boef (2008a,b, 2012b), I show that in the Dutch speaking language area there are basically six patterns (henceforth *systems*) of long-distance relativization, when focusing on the distribution of the elements *die* and *dat* in subject and object RCs with the common gender human antecedent *man*. These six systems are illustrated in table 4.9.[56] This table shows for each of the six systems of long relativization the corresponding system of short relativization. The numbers in the leftmost column indicate the number of locations (out of 203) in which a given system is found. When there are two numbers, the last number indicates the number of locations in which the system of long relativization is attested, and the first number indicates for how many of these long relativization systems the given short relativization pattern is attested. Sometimes not all locations that show the given pattern for long relativization thus also show the given pattern for short relativization, namely in the case of systems I and IV. The rightmost two columns in table 4.9 give the patterns of long-distance

[54] Parts of this section also appear in Boef (2008a), Boef (2008b) and Boef (2012b).

[55] See appendix A to chapter 2 for an evaluation of the analysis by Koopman and Sportiche (2008).

[56] The data in this section are drawn from the SAND corpus (unless indicated otherwise). As was already mentioned in chapter 1, I only make use of the material that was gathered in the oral and, when available, telephonic interviews (cf. section 1.3.1). Since I reanalyzed the relevant original data as published in SAND1 (Barbiers et al. 2005), the numbers and maps presented in this section sometimes differ (albeit slightly) from those published.

relativization for subjects and objects respectively. The way these columns should be read is as follows: the first of the two elements that each cell contains introduces the higher clause of a long-distance RC, whereas the second element introduces the lower (most deeply) embedded clause of a long-distance RC, e.g. *die-dat* thus stands for *de man* $[_{CP}$ ***die*** ... $[_{CP}$ ***dat*** ...]].

Table 4.9: Six systems of (long-distance) relativization in Dutch

n=203 (antecedent *man*)	short relativization		long relativization	
	subject	object	subject	object
system I (46/47)	die	die	die-dat	die-dat
system II (20)	die	dat	dat-die	dat-dat
system III (19)	die	die	die-die	die-dat
system IV (10/16)	dat	dat	dat-dat	dat-dat
system V (9)[57]	die	die	die-die	die-die
system VI (7)	die	die	dat-die	dat-die

The geographic distribution of these six systems is depicted on map 4.1.

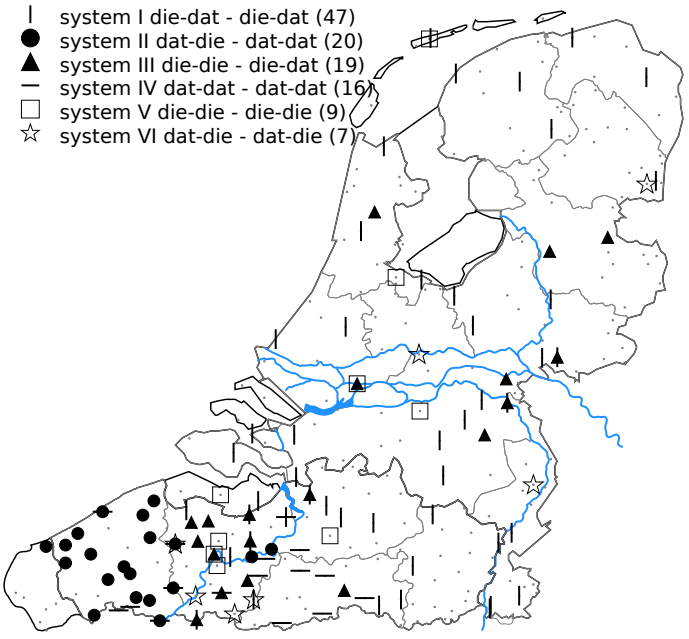


Figure 4.1: Six systems of long-distance relativization in Dutch

[57] As mentioned before, the numbers in this table are based on the SAND data. Subsequent research (the MPQ studies) has shown that doubling – the *die-die* pattern in table 4.9 – is way more widespread than represented here (cf. chapter 2).

On the basis of the observation that of all logically possible combinations of long-distance subject and object RCs,[58] only the six patterns in table 1 are attested, the following descriptive generalizations are formulated. Generalizations B and C are most relevant in the present context as they concern the two systems that show a s/o asymmetry (systems II and III).

(362) **A** long-distance relativization without a s/o asymmetry shows all logically possible variants: *die-dat*$_{SUBJ}$*-die-dat*$_{OBJ}$ (I), *die-die*$_{SUBJ}$*-die-die*$_{OBJ}$ (V), *dat-die*$_{SUBJ}$*-dat-die*$_{OBJ}$ (VI), and *dat-dat*$_{SUBJ}$*-dat-dat*$_{OBJ}$ (IV)

B a s/o asymmetry can only appear in the CP containing the extraction site (i.e. a s/o asymmetry in the higher clause is (almost) never attested)[59]

C in case of a s/o asymmetry (in the most deeply embedded clause), *die* occurs with subject extraction (= special *die*), whereas *dat* occurs with object extraction

Recall that in chapter 2, I gave an account of system I (Standard Dutch), and system V (doubling). I started from the assumption that doubling and non-doubling varieties have the same underlying syntax: successive-cyclic movement of the pronoun via the intermediate SpecCP. In Standard Dutch, only the head of the chain of the moved pronoun is spelled out, and in the lower clause the declarative complementizer shows up. Doubling, I argued, is the result of spelling out multiple copies, namely the head of the chain of the moved pronoun and the intermediate copy of the pronoun in SpecCP (see chapter 2 for details). These two derivations are illustrated in (363). Notice that both systems make use of 'relative pronoun' *die* for subject *and* object relativization, so the attested pattern for short relativization is exactly as it is predicted to be: pronoun *die* introduces both short subject and short object RCs (see table 4.9).

(363) a. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ ~~pronoun~~$_1$ ***dat*** ... ~~pronoun~~$_1$ ...]]
Standard Dutch, system I

b. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun~~$_1$ ...]]
colloquial Dutch, system V

I will show in section 4.5.3 that the existence of system IV follows naturally from the analysis. As for the pattern of system VI, the most straightforward analysis – that is in line with my analysis of doubling as discussed in chapter 2 – is to assume that it involves subextraction of the operator without spelling out the subextracted operator in the higher clause. As a result, the declarative

[58] With four possible patterns for long-distance subject RCs (*die-dat*, *die-die*, *dat-die* and *dat-dat*) and four possible patterns for long-distance object RCs (*die-dat*, *die-die*, *dat-die* and *dat-dat*), we expect 4^2 (=16) logically possible patterns of long-distance relativization.

[59] Cf. Haegeman (1983:84) who observes that the "*die-dat* alternation only occurs in the lowest Comp" in RCs in the dialect spoken in Lapscheure (West-Flanders).

complementizer *dat* instead of a pronoun will surface in the higher clause (because the left periphery in Dutch embedded clauses needs to be overt), as is abstractly illustrated in (364).

(364) $[_{CP}$ operator$_1$ ***dat*** ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun~~$_1$ ...]]
system VI

That this analysis of system VI might be on the right track is evidenced by the fact that system VI mainly occurs in an area that is surrounded by varieties that seem to make use of an empty operator when forming RCs (system II and system IV, cf. *infra*) and at the same time by varieties that use a pronoun to introduce RCs (system III and V, cf. *infra*). This might thus explain why system VI makes use of an empty operator (in the higher clause of long-distance RCs) as well as of a pronoun (in the lower clause of long-distance RCs, and in short RCs, cf. table 4.9).

In this section, I propose an account for the two systems of relativization that show a s/o asymmetry (systems II and III). Put differently, I try to provide an explanation for the two relevant generalizations in (362), i.e. generalizations B and C. The patterns of systems II and III are illustrated by the examples in (365) (=(55)) and (366) (=(56)) respectively.[60]

(365) a. Da s de vent **da** k peizen **die** da graptje verteld eet.
that is the man that I think *die* that joke told has
'That is the man who I think told that joke.'

b. Da s de vent **da** k peizen **da-n** ze geroepen en.
that is the man that I think that-3PL they called have
'That is the man who I think they have called'.
[Brugge Dutch, West-Flemish (system II)]

(366) a. Da s de man **die** k peize **die** t verhaal verteld ee.
that is the man RP I think *die* the story told has
'That is the man who I think told the story.'

b. Da s de man **die** k peize **da-n** ze geropen en.
that is the man RP I think that-3PL they called have
'That is the man who I think they have called.'
[Gent Dutch, East-Flemish (system III)]

In long-distance relativization in system II, the RC is introduced by *dat* and the finite embedded clause is introduced by *die* in case of subject extraction

[60]The West-Flemish data on special *die* (in (365)) are reminiscent of the well known *que/qui* alternation in French, i.e. *die* seems to have the same distribution as *qui* in long-distance relativization: it appears only in the left periphery of RCs from which the subject is extracted, cf. Bennis and Haegeman (1984:36). However, the distribution of special *die* across clause types is more restricted than the distribution of special *qui*: it only occurs in long-distance RCs (cf. *infra*, section 4.5.3). See also appendix A to chapter 2, in which it was shown that the parallel between special *qui* and special *die* is illusory: the two constructions differ from each other significantly.

and by *dat* in case of object extraction. This asymmetry is also attested with short relativization (see table 4.9).[61] System II is attested in 20 locations and it is found almost exclusively in West-Flanders, as can be seen on map 4.1. In system III, the higher clause in long-distance relativization is introduced by *die* and the lower clause is introduced by *die* in case of subject extraction and by *dat* in case of object extraction. This asymmetry is, however, not attested with short relativization, in which case *die* introduces both subject and object RCs.[62] System III is attested 19 times and its geographic distribution is somewhat less clear than that of system II. It is attested in the main part of East-Flanders and in several locations in the Netherlands: some isolated occurrences and a small cluster of attestations near the city of Nijmegen in the east of the Netherlands (cf. map 4.1).

The variable under study, i.e. the presence of a s/o asymmetry in the CP that contains the extraction site of movement, *covaries* with another variable in Dutch, namely the presence of *complementizer agreement* (CA).[63] Examples of CA can be found in (365b) and (366b): the complementizer in the lower clause agrees with the 3rd person plural subject *ze* 'they' in that clause, giving rise to the form *da-n*. Map 4.2 shows the geographic distribution of CA in the Dutch speaking language area. More specifically, this map shows for each location which members of the person and number paradigm show CA.

[61] As observed by Haegeman (1983) and Bennis and Haegeman (1984), West-Flemish – more specifically, the dialect of Lapscheure – has the option of using *dat* instead of special *die*. Put differently, the dialect of Lapscheure seems to make use of system II as well as of system IV. However, this pattern is not attested in the SAND corpus: only in 4 locations (out of the 20 locations in which system II is attested) can *dat* be used as an alternative for special *die* in short subject relatives (Oostkerke, Aalter, Berlare, Moerzeke), and of those 4 locations, there is only 1 location in which long-distance subject relativization can have the pattern *dat-dat* (Aalter).

[62] Actually, the data for system III are less clear than illustrated here. Of the 19 locations that make use of system III, all 19 locations make use of *die* in short subject and object relatives. But of these 19 locations there are also 10 locations that, in addition to *die*, can also use *dat* in short object relatives (and sometimes there are even more alternatives possible for both object and subject relatives). However, in what follows, I will assume that the pattern for system III as given in table 4.9 is in fact the core pattern, that is, a system that shows a s/o asymmetry with long-distance relativization, but no asymmetry with short relativization. The correctness of this assumption is corroborated by the observation that of the few locations that only make use of system III, the majority shows the pattern in table 4.9 (this claim is based on the data from the oral interviews in the SAND corpus).

[63] The insight that the presence of special *die* in the relevant varieties is somehow related to the presence of CA, was already noted by Bennis and Haegeman (1984). Their theory about special *die* is however completely different from the one proposed in this section: in their theory special *die* is a relative pronoun, not a complementizer. By virtue of CA may the pronoun in subject relatives be stranded in the lower CP domain, while the inflected complementizer moves to the higher CP domain. See section 4.5.4.1 for details.

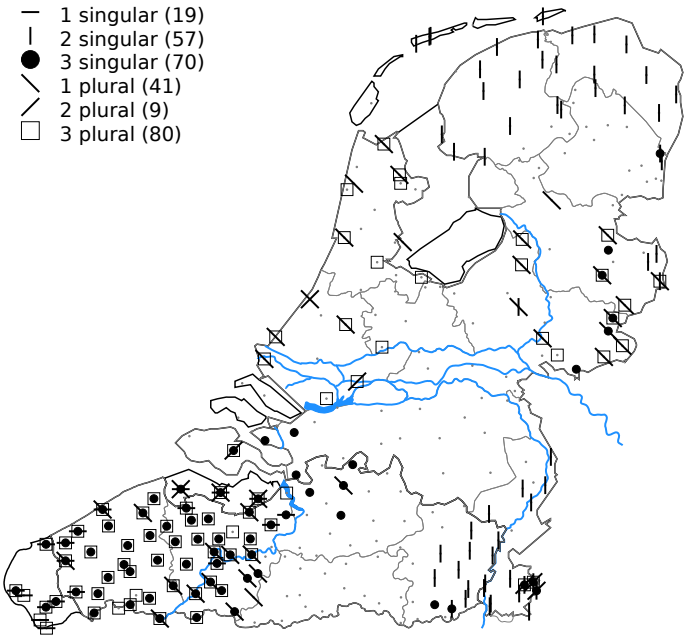


Figure 4.2: Dutch complementizer agreement (SAND1: Barbiers et al. 2005:21)

Comparing map 4.2 with map 4.1, we see that the locations that show a s/o asymmetry in long-distance RCs roughly form a subset of the locations that show CA. This is illustrated on map 4.3. Specifically, the geographic distribution of systems that show a s/o asymmetry (systems II and III) shows a significant positive correlation with the geographic distribution of CA.[64]

[64]The two variables, (i) +/- subject/object asymmetry, and (ii) +/- complementizer agreement (as scale variable), show a positive correlation (r=0.363 with n=203). This correlation is significant at the $p<0.01$ level (one-sided; p=0.00).

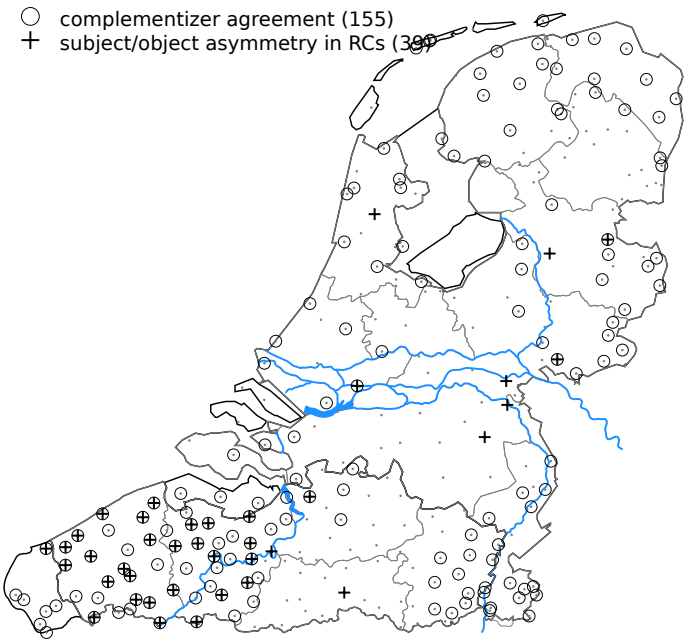


Figure 4.3: Correlation CA and subject/asymmetry in RCs

On the basis of this correlation, the following generalization is put forward. Notice that this generalization expresses a *one-way implication*, i.e. locations in which a s/o asymmetry in RCs is attested often also show CA, but it is not true that locations in which CA is attested also always exhibit a s/o-asymmetry in RCs.

(367) **D** the presence of a s/o asymmetry (special *die*) in long-distance RCs correlates with the presence of complementizer agreement

A similar claim has been made by Mayr (2010) for Bavarian (cf. also Rizzi 1990 for French).

4.5.2 Analysis: special *die* as a complementizer

Mayr (2010) argues for a treatment of long-distance subject extraction in Bavarian in terms of the relation with CA. More specifically, Mayr argues that CA licenses extraction of subjects. His proposal forms the starting point for my analysis of s/o asymmetries in (long-distance) RCs in varieties of Dutch, and will be discussed in the next subsection.

4.5.2.1 Mayr (2010)

In contrast to English, *that*-trace phenomena are absent in Bavarian, i.e. both subjects and objects are free to extract independently of the presence of a complementizer. In fact, extraction across a non-overt complementizer is never allowed (the same holds for Dutch, cf. chapter 2). This contrast is illustrated in (368) and (369).

(368) a. Who$_1$ does Mary believe [$_{CP}$ t_1 (*that) t_1 left Anna]?
b. Who$_1$ does Mary believe [$_{CP}$ t_1 (that) John left t_1]?
[English, Mayr 2010:117]

(369) a. Wea$_1$ hot da Michl gsogt [t_1 dass t_1 a Biachl kafft hot]?
who has the Michael said that a book bought has
'Who did Michael say bought a book?'
b. Wos$_1$ hot da Michl gsogt [t_1 dass d'Maria t_1 kafft hot]?
what has the Michael said that the Mary bought has
'What did Michael say that Mary bought?'
[Bavarian, Mayr 2010:117]

Mayr furthermore observes that Bavarian exhibits CA, as exemplified in (370). CA in Bavarian does not appear overtly on all persons: only second person singular and second person plural exhibit overt agreement morphology on C^0.

(370) a. Da Franz fragt [ob-**st** du morgen in d'Schui geh-st].
the Frank asks if.2SG you.2SG tomorrow in the school go.2SG
'Frank asks, if you(sg) will go to school tomorrow.'
b. Da Franz fragt [ob-**s** es morgen in d'Schui geh-ts].
the Frank asks if.2PL you.2PL tomorrow in the school go.2PL
'Frank asks, if you(pl) will go to school tomorrow.'
[Mayr 2010:120]

On the basis of the Bavarian data in (371)-(373), Mayr shows that CA correlates with extraction of subjects in Bavarian. The contrast in (371) shows that a subject can be extracted from the embedded clause only when there is CA. In contrast, the sentence in (372) shows that CA is not fully obligatory for all speakers when there is no subject extraction, indicating that the obligatory presence of CA in (371) is indeed related to subject extraction. Finally, the sentences in (373) show that in contrast to subjects, objects can freely extract, independently from CA (373a), and that objects in fact cannot agree with the local complementizer (373b).

(371) a. [Es Kinda]$_1$ hot da Hauns gfrogt [t_1 ob-s t_1 hamkummts].
you children has the John asked if.2PL home come
'John asked if you children will come home.'

b. * [Es Kinda]$_1$ hot da Hauns gfrogt [t_1 ob-Ø t_1 hamkummts].
you children has the John asked if.Ø home come
[Mayr 2010:121]

(372) % Da Hauns hot gfrogt [ob-Ø es Kinda hamkummts].
the John has asked if.Ø you children home come
'John asked, if you children will come home.' [Mayr 2010:121]

(373) a. [Die Bauan]$_1$ hot da Hans gefrogt [t_1 <ob-s/ob-Ø> es Kölna
the farmers has the John asked if.2PL/if.Ø you waiters
endlich t_1 bedients.
finally serve
'John asked if you waiters will finally serve the farmers.'

b. * [Eich Bauan]$_1$ hot da Hans gfrogt [t_1 ob-s die Kölna endlich
you farmers has the John asked if.2PL the waiters finally
t_1 bedienan.
serve
[Mayr 2010:122-123]

Traditionally, subject/non-subject asymmetries were explained by appealing to the *Empty Category Principle* (ECP; e.g. Chomsky 1981, Rizzi 1990). This principle states, simply put, that traces need to be licensed (*proper government*). Whereas traces of moved objects are automatically licensed by the selecting verb (*theta government*), licensing of subject traces requires an additional mechanism. More specifically, subject traces can only be licensed if they are locally bound by the moved constituent or its trace in the COMP domain (*antecedent government*).

Besides the observation that the ECP cannot properly explain the lack of *that*-trace effects (as in (369)), with the introduction of Minimalism, the ECP – specifically designed to regulate the distribution of traces – was abandoned. Mayr (2010) basically proposes a reformulation of the ECP in Minimalist terms, and argues that subject extraction is licensed by CA. I will briefly discuss his proposal in what follows.

Assuming that CA licenses subject extraction, the question that needs to be answered now is: what is the difference between extraction from SpecTP and extraction from the complement position of V? (Mayr 2010:118) Mayr argues that s/o asymmetries are a consequence of the manner in which Merge proceeds. More precisely, the fact that subjects are merged later in the derivation than objects gives rise to s/o asymmetries. The definition of *Internal Merge*, i.e. attraction of a goal by a probe, is given in (374).

(374) Internal Merge (IM)
IM at derivational stage Σ_i applies to nodes on the same projection line as the head H probing under c-command, thus a node formed at stage $\Sigma_{i-1}, \ldots, \Sigma_1$, where i $>$ 1. [Mayr 2010:131]

(375) Projection line (formulated in terms of *selection*)
X and Y are on the same projection line, iff the head X selects for YP. If the head Y selects for ZP, then by transitivity X and Z are on the same projection line. [Mayr 2010:131]

Given these definitions, it should be clear that objects can always undergo movement, because they are on the same projection line as a higher probing head. However, when an object has moved to the specifier position of a designated head, the question is how it can move any further, since a specifier is not on the projection line of a higher probing head (i.e. it is not selected for by that particular head). Mayr assumes that objects can always undergo movement, because they are on the projection line in their *External Merge* position. That is to say, since objects are merged onto the projection line when they are merged into the derivation, they can always undergo movement. Subjects, on the other hand, are never on the same projection line as a given head, because they are not selected for by any head. Therefore, given the definitions in (374)-(375), subjects cannot acts as goals, i.e. they can never be probed. In order for subjects to be probed, Mayr proposes (376), according to which a subject can connect to the projection line and be extracted by virtue of agreeing in ϕ-features with a head.

(376) The role of ϕ-features
Agreement between DP in SpecHP and H in ϕ-features connects an element to the projection line, of which H is a part. [Mayr 2010:132]

A subject in SpecTP can thus be extracted by virtue of agreeing in ϕ-features with T^0.[65] However, in order to derive that it is agreement with C^0 that licenses long-distance subject extraction, Mayr proposes (377), which is modeled on Chomsky's (2000) *Phase Impenetrability Condition* (PIC). The condition in (377) ensures that a subject in intermediate SpecCP position may undergo long-distance extraction only if it agrees in ϕ-features with the local C head. More specifically, when a subject has moved to the CP domain, its ϕ-agreement relation with T^0 is no longer visible and as a consequence, another ϕ-agreement relation with C^0 is established.[66]

[65]Something additional needs to be said about subjects of small clauses (or secondary predicates), which are subject-predicate structures lacking tense. It is unclear how the subject of a small clause – being in a specifier position – can extract. It cannot enter into a ϕ-agreement relation with the embedded C^0, as this C^0 can only agree with the local subject (*ik* in the Dutch example in (i)). I leave for future research the question of how exactly subjects of small clauses are connected to the projection line as a result of which they can extract. For now, I stipulate that subjects of small clauses (that are selected by a verb) are on the projection line in their external merge position, just like direct objects.

(i) Dat is de man [$_{CP}$ die Jan denkt [$_{CP}$ dat ik [$_{SC}$ _ intelligent] vind]].
that is the man RP Jan thinks that I intelligent find
'That is the man who Jan thinks I consider intelligent.'

[66]The ϕ-agreement relation between the subject in SpecCP and the local C head is visible because this relation is not contained in the *complement* of C^0 (cf. (377)), in which

(377) Deletion of ϕ-relations at Spell-Out
After S-O induced by a phase head P (e.g. C) ϕ-relations between a head and a DP in the complement of P are invisible for computation at the next phase. [Mayr 2010:133]

Assuming that selectional requirements need to be visible at all stages of the derivation, the following principle ensures that once the subject has entered into a ϕ-agreement relation with the local C^0, it is on the projection line once and forever (similar to objects) and need not enter into further ϕ-agreement relations.

(378) Secondary selection
If V selects for a CP with a DP in SpecCP which agrees with the head of this CP in ϕ-features, then the DP behaves as if it were directly selected by V. [Mayr 2010:137]

Notice that the principle in (378) predicts that obligatory CA – more generally, a s/o asymmetry – is always only found in the clause that contains the extraction site of the moved subject.[67] This is in line with generalization B in (362), which states that a s/o asymmetry in long-distance RCs is (almost) only attested in the most deeply embedded clause.[68]

4.5.2.2 The proposal

In order to account for the patterns of long-distance relativization in varieties of Dutch that show a s/o asymmetry, I follow Mayr's (2010) intuition that agreement with C^0 in the CP domain that contains the extraction site licenses subject extraction. More specifically, I will argue that special *die* in the varieties under discussion is in fact a specific manifestation of this agreement relation:

---

containment is defined as in (i).

(i) A ϕ-relation is contained in the complement of a phase head P if it holds between a head and a DP, both in the complement of P. [Mayr 2010:133]

[67] It is worth pointing out that the phenomenon of CA as discussed in the main text is different from the famous Irish CA. First, in long-distance extraction configurations in Irish, the complementizer of *each clause* that contains the trace of the extracted phrase shows agreement with that phrase, as illustrated in (i) for a long-distance RC (data like these are often cited as a strong argument in favor of successive-cyclic A-bar movement). Second, the Irish type of CA is not restricted to subject extraction, as can be seen in (i) where an object is extracted.

(i) an rud [**a** $C_{[+WH]}$ shíl mé [**a** $C_{[+WH]}$ dúirt tú [**a** $C_{[+WH]}$ dhéanfá]]]
the thing thought I said you do.COND.2SG
'the thing that I thought you said you would do' [McCloskey 1990:207]

[68] For English, Mayr (2010) assumes that in case of subject extraction there is no underlying C (hence the obligatory absence of the complementizer): the subject extracts directly from SpecTP (cf. Rizzi and Shlonsky 2007 for a similar proposal).

special *die* is an agreeing form of the complementizer *dat* (in line with traditional analyses of the French *que/qui* alternation, cf. Rizzi 1990) that shows up in RCs with a common gender antecedent.[69]

Claiming that special *die* is a form variant of the agreeing complementizer *dat* raises the issue of *morphological plausibility*. That is, although the varieties under discussion show CA, this agreement does not manifest itself as a *dat-die* alternation, but rather, as an agreement morpheme on complementizer *da* (e.g. (365b), (366b)). Put differently, the question that arises is why 'regular' CA is not enough to license subject extraction (as it is in Bavarian, cf. *supra*). In this respect, it is worth noting that there are in fact varieties of Dutch that act like Bavarian: they do not show s/o asymmetries in the sense that they do not feature special *die*, but they do obligatorily have agreement between the complementizer of the most deeply embedded clause and the extracted subject, as illustrated in (379).[70] The group of varieties that makes use of special *die* is thus a subpart of all varieties that show overt CA in the context of (subject) relativization.

(379) de mensen da-n ze zeggen <da-n/*da-Ø> da gekocht een
the people that-3PL they say that-3PL/that.Ø that bought have
'the people who they say have bought that'

From a cross-linguistic perspective, the proposed analysis of *die* in RCs as a special complementizer is not implausible as there are many more languages that have a special form of the complementizer that is only used in RCs (see de Vries 2002 for a comprehensive overview), like Icelandic, Slovenian (cf. (233) in section 3.4.4.4) and Tyrolean (cf. Alber 2008). Furthermore, there is even another language, namely Turkish, that displays a special morphology in subject RCs that is different from the morphology in object RCs or in embeddings

[69] In some West-Flemish locations, special *die* appears with an obligatory additional agreement affix when the antecedent is not third person singular, as illustrated in the following sentence for a 3rd person plural antecedent.

(i) de mensen da-n ze zeggen <**die-n/*die**> da gekocht een
the people that-3PL they say *die*-?/*die.Ø* that bought have
'the people who they say have bought that'

The question immediately arises as to why there is such additional CA if special *die* is an agreeing form of the complementizer itself. If *-n* in (i) indeed shows additional CA, similar to what we find on complementizer *da*, we expect to find *die-t* instead of *die* in case the relative subject is third person singular; *-t* being the agreement affix that shows up on the complementizer with third person singular subjects (in all Dutch locations that have CA marking third person singular, cf. SAND1 data (Barbiers et al. 2005:19)). However, this prediction is not borne out. Most likely, the *-n* affix in examples like (i) is the phonological realization of a feature that is different from the features that are phonologically realized by *die*. Most likely, *-n* is a number marker marking plural, while *die* only expresses common gender. Further research into the whole paradigm of 'additional' CA affixes on special *die* is necessary to see if data as in (i) are representative for the whole language area that makes use of special *die*.

[70] Dutch varieties might be different from Bavarian in having *obligatory* CA, in which case it becomes very hard to see whether or not CA licenses subject extraction, as there will always be CA independently from whether or not subject extraction has taken place.

in general. Interestingly, Kornfilt (2009) also attributes this difference to CA effects.

Given the observation that elements like Dutch *dat* or English *that* have a dual status – i.e. they are demonstrative/relative pronouns *and* may act as complementizers (cf. section 4.3.2) – it should not come as a big surprise that *die* has a dual status as well: demonstrative/relative pronoun and complementizer. Assuming that *dat* developed from the demonstrative/relative pronoun *dat* (by a process of grammaticalization, cf. Roberts and Roussou 2003), it might be conceivable to think of special *die* as having developed from the demonstrative/relative pronoun *die* as well.[71] If special *die* indeed finds its origin in the relative pronoun *die*, the observation that special *die* has the form of a relative pronoun, and the fact that the distribution of special *die* is restricted to RCs, can be accounted for. More specifically, if special *die* were merely a mechanism to license subject extraction, special *die* would be predicted to show up in constructions different from RCs as well, for example in long-distance *wh*-Qs, as indicated in (380). In other words, varieties that make use of special *die* in (long-distance) subject RCs would be predicted to make use of the construction in (380a) when forming long-distance subject *wh*-Qs, while at the same time using (380b) for long-distance object *wh*-Qs. In (380a) the question is introduced by the *wh*-word *wie* 'who', and the embedded clause is introduced by special *die*, which is taken to be a complementizer spelled-out as *die* by virtue of agreement with (a copy of) the subject *wh*-phrase.

(380) a. **Wie** denk je **die** het verhaal verteld heeft?
who think you *die* the story told has
'Who do you think told the story?'

b. **Wie** denk je **dat** ze geroepen hebben?
who think you that they called have
'Who do you think they have called?'

However, special *die* does not seem to occur in long-distance *wh*-Qs, which strengthens the idea that special *die* is a derivative of relative pronoun *die*.[72] In fact, such an account of special *die* makes an even more specific prediction,

[71] It thus appears to be the case that pronouns can be 'reanalyzed' as heads under certain circumstances. Bayer and Brandner (2008) for example show that in certain dialects of German the simplex interrogative pronoun in *wh*-Qs seems to appear in the position of the complementizer, making doubly filled Comp impossible (cf. Westergaard and Vangsnes 2005, and Vangsnes 2005 for a similar observation in certain Norwegian dialects, e.g. the dialect of Tromsø).

[72] As already became clear in chapter 2, patterns like (380a) are attested in the Dutch speaking language area. However, this does not tell us anything, because the occurrence of (380a) is only interesting in relation to the pattern in the corresponding long-distance object *wh*-Q (380b): only in case the lower clause of the corresponding long-distance object *wh*-Q is introduced by the declarative complementizer *dat* can we be sure to be dealing with special *die* in the subject extraction case (although even then it still might be a case of *doubling*). Unfortunately, I was unable to test the prediction regarding the distribution of special *die* on the basis of the SAND corpus, as in the SAND project sentences like (380a) were only questioned with a *wh*-object, and not with *wh*-subjects. Also the MPQ data were not very

namely that only in RCs that independently can be introduced by relative pronoun *die* – i.e. RCs with a common gender antecedent (*die* being specified for gender as [common], cf. section 2.5.2.4) – can special *die* occur. Unfortunately, (long-distance) RCs with neuter antecedents were not systematically tested in the SAND project nor in the MPQ, so this prediction cannot be checked on the basis of those corpora. Interestingly, Liliane Haegeman (p.c.) informs me that the use of special *die* in her dialect (Lapscheure, West-Flanders) is indeed related to gender: it may only occur with common antecedents, but not with neuter antecedents.[73]

Having established that it is not implausible to analyze special *die* as an agreeing variant of the complementizer, recall from table 4.9 that not all varieties of Dutch show a s/o asymmetry in (long-distance) relativization. Starting from the null hypothesis that states that whenever there is subject extraction, the subject needs to agree with the local complementizer (if CP is present, cf. Mayr 2010), it seems straightforward to assume that different varieties differ with respect to whether they spell out this agreement relation or not. And if so, how they spell out this agreement relation: by 'regular' CA or by special *die* (or by a combination of both, cf. footnote 69). This is formulated by the following micro-parameter.[74]

(381) micro-parameter 1: +/- spell out of agreement with C^0 (as *die*)

An important consequence of this proposal is that it implies that in case of subject extraction there always needs to be CA in Dutch, independently of whether or not this is spelled out (cf. Mayr 2010).[75] To account for the fact that the majority of Dutch varieties does not spell out this agreement relation as *die*, the most straightforward assumption is that not all varieties have *die* as a complementizer in their lexicon, in which case complementizer *dat* will be used as the default. This proposal can be formally implemented as follows: *die*

insightful into the matter, as there were no (or not many) MPQ informants in the part of the Dutch speaking language area in which special *die* is attested (according to the SAND data). However, as observed by Haegeman (1983:83) for the dialect of Lapscheure (West-Flemish) – and as indicated by all West-Flemish and East-Flemish informants I checked with personally – the distribution of special *die* is indeed restricted to RCs, i.e. special *die* does not occur in the most deeply embedded clause of long-distance subject *wh*-Qs. Needless to say, further research is required to settle the issue of the distribution of special *die* across clause types.

[73]Recall that in the dialect of Lapscheure, the use of special *die* is never obligatory (*dat* may be used in special *die* contexts as well, cf. footnote 61). It would therefore be particularly interesting to check the prediction that the distribution of special *die* is restricted to subject RCs with a common gender antecedent with speakers for whom special *die* appears obligatorily in subject RCs.

[74]I remain agnostic about the exact status of parameters in the model of grammar. Parameters like the one in (381) can easily be reduced to differences in the Lexicon or Vocabulary (cf. section 1.2.1) – in line with Chomsky (1995) who argues that the lexicon is the locus of (micro)variation (cf. section 1.2.2).

[75]Interestingly, Postma (2006) makes a similar claim on different grounds (a study on number neutralization of the subject pronoun *ze* 'she'/'they' in Dutch). He argues that although the (ϕ-)agreement features in C^0 are always (abstractly) present (in asymmetrical V2 languages), they are not spelled out in all varieties.

as a complementizer is more specific than *dat* (like pronoun *die*, complementizer *die* is specified for gender as [common], cf. *supra*), and will therefore always be inserted in case of competition between *dat* and *die*.

At this point we can give an explanation for generalizations B, C and D, which are repeated here for convenience in (382).

(382)
- **B** a s/o asymmetry can only appear in the CP containing the extraction site (i.e. a s/o asymmetry in the higher clause is (almost) never attested)
- **C** in case of a s/o asymmetry (in the most deeply embedded clause), *die* occurs with subject extraction (= special *die*), whereas *dat* occurs with object extraction
- **D** the presence of a s/o asymmetry (special *die*) in long-distance RCs correlates with the presence of complementizer agreement

As already mentioned, generalization B can be accounted for by Mayr's mechanism of Secondary Selection (cf. *supra* (378)): once the subject has entered into an agreement relation with the most deeply embedded C^0, it acts as though it is directly selected for by the verb (i.e. it is on the projection line once and forever) and therefore it can be extracted and need not enter into further agreement relations with higher (C^0) heads. Generalization C can be accounted for by the assumptions that special *die* is an agreeing form of complementizer *dat*, and that only the subject, not the object, needs to agree with the local complementizer in order to be on the projection line and get extracted, i.e. special *die* only occurs with subject extraction. Finally, when special *die* is taken to be an agreeing variant of the complementizer, generalization D is explained: specifically varieties that show agreement morphology on the complementizer in other contexts (see map 4.2), also spell out the agreement relation with the local complementizer in RCs.

Although the proposed analysis can account for generalizations B, C and D, the difference between systems II and III is not yet explained. Put differently, micro-parameter 1 only distinguishes varieties that show a s/o asymmetry (systems II and III) from varieties that do not show a s/o asymmetry (systems I, IV, V and VI). As the proposed analysis only has something to say about what is going on in the most deeply embedded clause of long-distance RCs: die_C with subjects and dat_C with objects, there is still the need for an explanation of the patterns in the higher clause of long-distance RCs: *dat* in system II and *die* in system III for both subjects and objects. In order to answer the question of what is going on in the higher clause, the patterns for short relativization, repeated here in table 4.10, become important.

Table 4.10: Systems that show a subject/object asymmetry

n=203	short relativization		long relativization	
	subject	object	subject	object
system II (20)	**die**	**dat**	dat-die	dat-dat
system III (19)	**die**	**die**	die-die	die-dat

Assuming for the moment that not only Standard Dutch RCs (cf. *supra*), but RCs in all varieties of Dutch need to be introduced by a relative pronoun, we predict that in a given location the pattern of short relativization is identical to the pattern in the higher clause of long relativization. Whereas that assumption makes the right prediction for the pattern of system III: *die* is used with short relativization and in the higher clause of long relativization (see *infra*), it makes the wrong prediction for system II: long-distance relativization in system II does not show an asymmetry in the higher clause (see table 4.10), yet short relativization does. This observation seems to force the conclusion that the assumption that all varieties of Dutch require a relative pronoun to introduce the RC is incorrect. The pattern of system II can then easily be explained by assuming that system II, in contrast to system III, does not make use of relative pronouns, i.e. system II uses an empty operator when forming RCs. Together with the observation that in all Dutch varieties there always needs to be at least one overt element in the complementizer domain, we have an explanation for why *dat* introduces the higher clause in long-distance RCs in system II. Because system II uses an empty operator instead of an overt relative pronoun and because there needs to be an element overtly present in the left periphery, the complementizer (*dat*) appears. Given the observation that short relativization in system II also shows a s/o asymmetry, it needs to be assumed that the relative pronoun or operator always enters into a ϕ-agreement relation with the most embedded C^0 in case of subject relativization, i.e. the ϕ-agreement relation with the C^0 in the CP domain that contains the extraction site is established in both short and long relativization.[76]

Whereas system II thus does not make use of overt relative pronouns, for varieties that show the pattern of system III, it is assumed that they make use of relative pronouns, namely *die* for both subjects and objects. The pattern for this system then follows naturally: because the RC is always introduced by the relative pronoun *die*, there is no asymmetry in short relativization (only the pronoun is visible, not the complementizer) or in the higher clause in long relativization.[77]

[76]Recall that according to Mayr, at the point the subject is in SpecCP, the agreement relation with T^0 has become inaccessible due to Spell Out. For the subject to be *extracted* from SpecCP, it needs to enter into a ϕ-agreement relation with C^0. Even though in short relativization the relative pronoun or operator does not extract from SpecCP, I assume that it obligatorily enters into an agreement relation with the local C^0.

[77]One could imagine that there are varieties that make use of system III and at the same time allow doubly filled Comp. Although such varieties are predicted to have the string die_{RP} die_{C} in the C domain of short subject RCs, such strings are not attested in the SAND

The difference between systems II and III can thus be attributed to the use or spell out of relative pronouns, as described by the following micro-parameter.

(383) micro-parameter 2: +/- spell out of relative pronouns

On the basis of the two proposed micro-parameters, the properties of systems II and III can be summarized as follows. Both systems spell out the agreement relation with the local complementizer in case of subject extraction (special *die*), but the systems differ with respect to whether they spell out relative pronouns or not (overt relative pronoun versus empty operator). This is summarized in table 4.11.

Table 4.11: Summary analysis system II and system III

	short relativization		long relativization	
	subject	object	subject	object
+ spell out C_{AGR} as *die* & – spell out RPs = system II	die_C	dat_C	dat_C-die_C	dat_C-dat_C
+ spell out C_{AGR} as *die* & + spell out RPs = system III	die_{RP}	die_{RP}	die_{RP}-die_C	die_{RP}-dat_C

4.5.3 Predictions and empirical support

The proposed analysis makes several predictions. First, if special *die* is an agreeing variant of the declarative complementizer *dat*, special *die* is predicted to be in complementary distribution with the declarative complementizer *dat*. More specifically, varieties that make use of special *die* are predicted to never allow the string *die dat* in contexts with special *die*. This prediction is largely borne out by the SAND data: 38 out of the 39 locations in which special *die* occurs in (long-distance) RCs do not show the string *die dat* in special *die* contexts.[78] This conclusion is further corroborated by the following observation. Normally,

corpus. A plausible explanation for this is that natural languages have the tendency to avoid the (accidental) repetition of identical morphemes within a particular syntactic environment (cf. Neeleman and van de Koot 2006). As a result of this, only one *die* in the string die_{RP} die_C in short subject RCs is maintained, but it is unclear which *die* this is. In table 4.11, I assumed it is relative pronoun *die* that is maintained in short subject RCs in system III.

[78] It is important to note that not all SAND informants were explicitly asked about the potential presence of a complementizer in addition to the element *die*. Therefore, the prediction that a complementizer cannot be present in special *die* contexts cannot be adequately tested on the basis of the SAND corpus. I tried to test the prediction in MPQ1-B, but those data are not very insightful either, as special *die* is not (or only very marginally) attested in that corpus in the first place (cf. section 2.3.1); the locations in which special *die* is attested as reported in the SAND corpus are (for the most part) not the same locations for which there are MPQ1-B data (cf. map 1.2 in section 1.3.2). Needless to say, further (empirical) research into the issue of doubly filled Comp in special *die* contexts is necessary.

in West-Flemish subordinate clauses the complementizer is always overt, independently of the presence of an element in SpecCP (Haegeman 1992:57). This is illustrated in (384)-(385).

(384) a. Kpeinzen da Valère a weg is.
I think that Valère already away is
'I think that Valère is already gone.'

b. * Kpeinzen Ø Valère a weg is.

(385) a. Kweten niet wannièr da Valère goa werekommen.
I know not when that Valère goes return
'I do not know when Valère is going to return.'

b. * Kweten niet wannièr Ø Valère goa werekommen.

[West-Flemish, Haegeman 1992:57]

However, in West-Flemish RCs, complementizer *dat* is never overt in the presence of a relative pronoun (cf. Haegeman 1983). Whereas this observation is puzzling under the assumption that special *die* is a relative pronoun, it follows straightforwardly from the analysis of special *die* as a complementizer. Under that assumption, the generalization that in West-Flemish subordinate clauses the complementizer always needs to be overt can be maintained, as special *die* itself is the complementizer.

Second, the proposed analysis predicts the existence of the four systems of relativization as given in table 4.12. This prediction is borne out: all four systems of relativization are attested within the SAND corpus. In fact, the two proposed micro-parameters generate four out of the six systems of (long-distance) relativization that were presented in section 4.5.1. In other words, in addition to the patterns of the systems that show a s/o asymmetry (systems II and III) the patterns of system I and system IV follow naturally from the proposed analysis. For an analysis of the two remaining systems of relativization (systems V and VI), see again section 4.5.1.

Table 4.12: Predicted patterns of long-distance relativization

	+ spell out agreement	- spell out agreement
- relative pronouns	dat_C-die_C, dat_C-dat_C (system II)	dat_C-dat_C, dat_C-dat_C (system IV)
+ relative pronouns ($\textit{die}_{SUBJ}$, $\textit{die}_{OBJ}$)	die_{RP}-die_C, die_{RP}-dat_C (system III)	die_{RP}-dat_C, die_{RP}-dat_C (system I)

In addition to this prediction regarding the *existence* of different systems of relativization, the analysis also makes a prediction with respect to the *non-existence* of certain systems: because the analysis only predicts the existence of the systems in table 4.12, at the same time the analysis predicts the non-existence of all other logically possible patterns of (long-distance) relativization

that show a s/o asymmetry. Table 4.13 shows that this prediction is borne out: all other systems that are ruled out by the proposed analysis (in the sense that they violate one or two of the generalizations on which the analysis is built) are not or only very marginally attested in the SAND corpus.[79]

Table 4.13: Excluded patterns of long-distance relativization

long subject	long object	gen. B	gen. C	# attestations in SAND corpus
die-dat	die-die	+	−	2
die-dat	dat-dat	−	+	5[80]
die-dat	dat-die	−	−	0
die-die	dat-dat	−	+	0
die-die	dat-die	−	+	0
dat-dat	die-dat	−	−	1
dat-dat	die-die	−	−	0
dat-dat	dat-die	+	−	0
dat-die	die-dat	−	−	0
dat-die	die-die	−	−	1

In sum, it can thus be stated that the proposed analysis correctly predicts the existence of *all* and *only* those patterns of long-distance relativization with a s/o asymmetry that are attested in the SAND corpus.

4.5.4 Alternative accounts

In appendix A to chapter 2, I discussed an alternative analysis of special *die* (and doubling) – namely the analysis by Koopman and Sportiche (2008) – which was shown to be inadequate on empirical as well as on theoretical grounds. In this section, I will briefly discuss another explicit analysis of special *die*, namely the analysis by Bennis and Haegeman (1984), who argue that special *die* is a relative pronoun. More specifically, in light of the discussion of pronouns

[79]The numbers in the rightmost column of this table indicate the number of locations in which the given combination of a long subject relative and a long object relative is attested. More specifically, this column shows in how many locations the particular pattern of long-distance relativization is used *for sure*; thus, only locations in which no other combinations of subject and object relatives are attested are taken up in this table.

[80]Although the pattern *die-dat*$_{SUBJ}$–*dat-dat*$_{OBJ}$ violates generalization B and should thus be unattested, it ís attested in 5 locations in the Dutch speaking language area. The occurrence of this pattern (albeit very marginal) might be explained by assuming that in these locations the agreement relation with the local complementizer in case of subject extraction is not spelled out (no special *die*) and that a different relative pronoun is used for subject relatives and object relatives: namely *die* and *dat* respectively (notice that micro-parameter 2 (+/- relative pronouns) does not say anything about the form of the relative pronouns). The observation that in 3 out of these 5 locations short subject RCs are introduced by *die* and short object RCs are introduced by *dat* suggests that this line of reasoning is on the right track.

and complementizers in the preceding sections, I will show that assuming that special *die* is a relative pronoun cannot be correct.

Besides specific alternative proposals regarding the nature of special *die*, there exist many more analyses of s/o asymmetries more generally (that do not rely on the *Empty Category Principle*). In particular the work by Rizzi and Shlonsky (2007) is important in this respect. This will be briefly discussed in subsection 4.5.4.2.

4.5.4.1 Bennis and Haegeman (1984)

Bennis and Haegeman (1984) (henceforth B&H) argue that special *die* in long-distance subject RCs is a relative pronoun that remains in the lower CP, whereas the complementizer of the lower CP moves up to the higher clause where it is deleted (cf. Bennis and Haegeman (1984:50ff.) for the precise workings of *deletion under identity* in the realm of the left periphery of RCs). The movement of the complementizer is assumed to be possible by virtue of CA: the inflected complementizer contains pronominal features and is related to the subject position by *superscripting*, as indicated in (386a). This approach immediately accounts for the lack of special *die* in long-distance object RCs: because there is no CA with the object, the relative pronoun cannot be stranded. The relative pronoun in object RCs thus successive-cyclically moves to the highest CP domain, where it is deleted under identity with the RC head. This is indicated in (386b).

(386) a. ... RC head $[_{CP}$ ~~**dat**~~$_2$ **dat** ... $[_{CP}$ **die**$_1$ t^i_2 $[_{IP}$ t^i_1 ...]]]
long-distance subject relativization

b. ... RC head $[_{CP}$ ~~**die**~~$_1$ **dat** ... $[_{CP}$ t_1 **dat** $[_{IP}$ ... t_1 ...]]]
long-distance object relativization

If we only take into account the pattern of system II (West-Flemish) – as B&H do – the relevant generalizations can also be explained by their analysis. First, the observation that the distribution of special *die* is restricted to RCs follows straightforwardly as special *die* is in fact a relative pronoun. Second, the correlation with CA (generalization D) is accounted for as the complementizer may only move and strand the relative pronoun by virtue of it agreeing with the local subject. The fact that special *die* only occurs with subjects (generalization C) is thus accounted for, because CA is only possible with subjects in West-Flemish. The fact that special *die* only occurs in the most deeply embedded clause (generalization B) is explained by the fact that the relative pronoun (*die*) can only be stranded in the clause in which it originated, as it is only in that clause that the complementizer can agree with the subject (and subsequently extract). Finally, the observation that DFC is unattested in special *die* contexts is accounted for as follows. Under the assumption that there is obligatory deletion in the CP domain in West-Flemish RCs (Bennis and Haegeman 1984:51), either the complementizer or the relative pronoun is targeted by deletion in *short* subject RCs (in short *object* RCs, only the relative pronoun can

be targeted by deletion, cf. Bennis and Haegeman 1984 for discussion). In *long* subject RCs, the complementizer moves out from the lower CP, thereby leaving the relative pronoun as the only overt element in the CP domain (hence the absence of DFC).

Besides some problems with the theoretical apparatus that the analysis makes use of (from the perspective of contemporary theorizing) – e.g. the crucial reliance on *indices* (in violation of the *Inclusiveness Condition*, Chomsky 1995:228) – the analysis of B&H cannot provide a unified account of systems II and III. As noted before, the analysis makes the right predictions when only the pattern of system II is taken into account. However, when taking into account the pattern of system III as well – which also features special *die* – the analysis makes the wrong predictions, leaving unexplained certain crucial properties of the pattern of system III. The reason for that is that in system III the element *die* occurs with subjects as well as with *objects*, in short relativization and in the higher clause of long relativization. More specifically, if special *die* were to be a stranded relative pronoun – as argued for by B&H – it is unclear how it can occur in the higher clause of long-distance relativization as well. Furthermore, the deletion under identity mechanism that B&H assume for system II needs to be adjusted to account for system III, thereby losing a unified account of systems that make use of special *die*. An analysis of special *die* in terms of it being a relative pronoun is thus not very well founded if we look beyond the West-Flemish pattern.[81]

4.5.4.2 Rizzi and Shlonsky (2007)

A recently formulated alternative account to the ECP is provided by Rizzi and Shlonsky (2007) (henceforth R&S). Their account is based on the following two ideas, presented in (387).

(387) a. **Criterial Freezing**: an element moved to a position dedicated to some scope-discourse interpretive property, a criterial position, is frozen in place (cf. Rizzi 2006, 2010)

b. The Classical EPP – clauses need subjects – is restated as the **Subject Criterion**

These formulations ensure that when a subject moves to the criterial subject position – the highest head within TP: $Subj^0$ – it is frozen in place. The difficulty of moving subjects is thus explained by appealing to Criterial Freezing. Since there is no such thing as an Object Criterion, the distribution of objects is

[81] It has been suggested to me several times that special *die* might be a manifestation of *wh*-agreement (similar to Irish CA, cf. footnote 67). This cannot explain why special *die* only occurs with subjects and not with objects (generalization C), why special *die* only occurs in the most deeply embedded clause (generalization B), and why the occurrence of special *die* is correlated with CA (generalization D). Furthermore, under a *wh*-agreement approach it remains unclear why special *die* (most likely) only occurs in RCs, and not in other (long-distance) A-bar dependencies, like *wh*-Qs.

not similarly constrained. Sentences in which the subject cannot extract (e.g. in English) now follow straightforwardly: the subject moves to the specifier of Subj^0, where it is frozen is place and not allowed to move any further by Criterial Freezing.

In order to account for the French *que/qui* alternation, R&S adopt the main insights of the analysis by Taraldsen (2001). Taraldsen assumes that *qui* is a contracted form, composed of *que* and the expletive element *i* (reminiscent of the Standard French expletive *il*). The basic idea is that in subject RCs in French, the expletive element *i* – which is merged as the finiteness head (Fin^0), just below F^0 which hosts *que*, and just above SubjP, i.e. [$_{FP}$ *que* [$_{FinP}$ *i* [$_{SubjP}$]]] – satisfies the Subject Criterion. Due to its nominal nature (i.e. *i* is merged with a set of ϕ-features), the expletive *i* in Fin^0 is assumed to satisfy the nominal requirement of Subj^0, as it is in a head-head relation with Subj^0. As a consequence, the thematic subject is free to move to the CP domain, i.e. the thematic subject can skip the criterial position. Put differently, the subject first moves to SpecFin – which is not criterial – to value the ϕ-features of *i*, after which it can move further to the CP domain. Simply put, the subject can freely extract only if *i* is present to satisfy the Subject Criterion, which gives rise to the appearance of *qui*. The impossibility of *qui* appearing in object RCs is explained by the fact that there is simply not enough room for the expletive *i* and the subject DP to occur in the same structure, i.e. they are mutually exclusive. For the precise implementation of this analysis, I refer the reader to Rizzi and Shlonsky (2007). For the present discussion, the above short outline will suffice. In the remainder of this section, I will show, mainly based on arguments by Sportiche (2008) and Mayr (2010), that R&S's analysis of s/o asymmetries is empirically and theoretically inadequate.

According to Sportiche (2008), R&S's proposal faces two serious problems. First, the assumptions that (i) special *qui* is composed of the heads *que* (F^0) and *i* (Fin^0), and that (ii) the existence of two heads is justified by the possibility of having material intervening between the two heads (e.g. adjuncts), predict the string *que X i* to exist. Such strings are never attested. Second, main clause *wh*-Qs like (388b) are predicted to be well-formed – just as in embedded *wh*-Qs the Subject Criterion is satisfied by *i*, opening up the possibility for the thematic *wh*-subject to move to the higher left periphery – contrary to fact.[82]

(388) a. Quel enfant est parti?
'Which child left?'

b. * [Quel enfant]$_1$ [$_F$ e] t_1 [$_{Fin}$ ***i***] [$_{Subj}$ t_1 ...

[French, Sportiche 2008:76]

To account for the ill-formedness of (388b), R&S assume that in French main clauses, a silent version of [$_{Fin}$ *i*] is allowed. Leaving aside the fact that this is

[82] Notice that these points of criticism regarding Rizzi and Shlonsky's (2007) approach also hold for Taraldsen's (2001) analysis of the *que/qui*-alternation.

merely an *ad hoc* solution to the problem at hand, naturally – as pointed out by Sportiche (2008) – this assumption faces the problem of overgeneration.

A third point of criticism regarding the account of R&S, as noted by Mayr (2010), has to do with the prediction that their analysis makes with respect to CA patterns in coordination structures. In Bavarian, there are two possible CA patterns for an in-situ coordinated subject that consists of a 2nd person singular subject and a 3rd person singular subject, namely 3rd person plural CA (389a) or 2nd person plural CA (389a) – only the latter results in *overt* CA. Interestingly, in case the coordinated subject extracts, as illustrated in (390), only the pattern with overt CA is possible. Whereas Mayr accounts for this contrast by assuming that whenever overt CA is possible, it should be exhibited to overtly indicate the licensing of subject extraction, the ungrammaticality of (390b) is unexpected by R&S's analysis. As it is irrelevant how the Subject Criterion is satisfied (as long as it is satisfied), i.e. by 2nd person plural or by 3rd person plural features, R&S predict (390b) to be as grammatical as (390a), *qoud non*.

(389) a. Da Hauns hot gfrogt [ob du und da Franz weggengan].
the John has asked if you.2SG and the Frank leave-3PL
'John asked if you and Frank will leave.'

b. Da Hauns hot gfrogt [ob-s du und da Franz weggeh-ts].
the John has asked if-2PL you.2SG and the Frank leave-2PL
'John asked if you and Frank will leave.'

(390) a. [Du und da Franz]$_1$ hot da Hauns gfrogt [t_1 ob-s t_1 weggeh-ts].
you.2SG and the Frank has the John asked if-2PL leave-2PL
'John asked if you and Frank will leave.'

b. * [Du und da Franz]$_1$ hot da Hauns gfrogt [t_1 ob t_1 weggengan].
you.2SG and the Frank has the John asked if leave-3PL
[Mayr 2010:122]

Mayr furthermore observes that the analysis makes the wrong empirical predictions regarding CA constructions. R&S assume that the position that deals with the Subject Criterion (Subj^0) is different from the position that deals with agreement (Agr^0) – in which Agr^0 is directly below Subj^0 in the clausal structure – and that in expletive constructions, the subject moves to SpecAgrP, whereas the expletive moves to SpecSubjP to satisfy the Subject Criterion. However, the English construction in (391) shows that this assumption cannot be right: there does not seem to be any subject movement as the subject follows the low adverb *often*, which marks the edge of *v*P.

(391) There is often a room available. [Mayr 2010:141]

The grammaticality of (391) leads Mayr to conclude that the agreement relation between the subject and Agr^0 in expletive constructions can be a *long-distance* relation. This conclusion makes a clear prediction with respect to CA constructions. Under R&S's account, CA is analyzed on a par with expletive constructions in the sense that in case of CA, the ϕ-features on the complementizer satisfy the nominal requirement of $Subj^0$ and – just like in expletive constructions as in (391) – the subject stays *in-situ* (i.e. it does not move to SpecAgrP) from where it agrees long-distance with Fin^0. Now an interesting test case is VP-fronting in German: given the above assumptions, a VP containing the subject should be able to be fronted in case of CA, as the ϕ-features on C^0 may satisfy the Subject Criterion. However, as illustrated in (392b), this prediction is not borne out, leading to the conclusion that ϕ-features on the complementizer cannot satisfy the Subject Criterion. Put differently, the subject does not skip SpecSubjP. Mayr (2010:142) concludes that any account "where the ϕ-features of C^0/Fin^0 satisfy the EPP is unfeasible".

(392) a. $[_{vP}$ A Buach kafft$]_1$ hot da Hauns gfrogt $[_{CP}$ ob-s es
a book bought has the John asked if-2PL you.2PL
t_1 hobts].
have
'John asked whether you(pl) bought a book.'

b. * $[_{vP}$ Es a Buach kafft$]_1$ hot da Hauns gfrogt $[_{CP}$ t_1 ob-s
you a book bought has the John asked if-2PL
es t_1 hobts].
you.2PL have [Mayr 2010:141-142]

From the above, it is clear that the theory of Rizzi and Shlonsky (2007) faces some serious conceptual/theoretical as well as empirical shortcomings. It should be noted, however, that their analysis does make the right predictions regarding the Dutch systems of relativization that show a s/o asymmetry. It follows from R&S's proposal that special *die* is only attested in the most deeply embedded clause of long-distance subject RCs, because the relativized constituent does not have to move through the SpecSubjP in the higher clause, but only in the lower clause. That is to say, only in the lower clause we need an expletive-like element that satisfies the Subject Criterion (the Subject Criterion in the higher clause is satisfied by the local subject). Moreover, under the assumption that special *die* is a contracted form that is composed of the *complementizer* and another element in SpecSubjP[83] – instead of a (weak) relative pronoun (as in Bennis and Haegeman 1984, Koopman and Sportiche 2008) – the analysis of R&S makes more or less the same predictions that my analysis does.

[83] I will not go into the issue of whether or not it could be plausibly argued that special *die* is a contracted form consisting of the complementizer together with some element in SpecSubjP.

4.6 Conclusion

Many languages feature relative pronouns that are taken from another pronoun paradigm, like the demonstrative or the interrogative pronoun paradigms. Relative pronouns thus often may have more than one function and may appear in more than one configuration (multipurpose pronouns). The aim of the first part of this chapter was to show that such multipurpose pronouns have a single *underspecified* lexical entry, as a result of which they may appear in more than one syntactic environment (null hypothesis). For example, I argued – following a proposal by Rooryck (2003) – that distal demonstratives are underspecified for location, whereas proximal demonstratives are specified for location, and that it is exactly this underspecification for location that makes distal demonstratives suitable to function as relative pronouns.

Just like relative pronouns, finite declarative complementizers in many languages are taken from the demonstrative or interrogative pronoun paradigms as well. The Dutch complementizer *dat* 'that' corresponds to the neuter gender distal demonstrative. Without taking the complementizer to be identical to this pronoun (i.e. they do not have the same lexical entry), but assuming they are diachronically related nonetheless, I argued – in correspondence with the null hypothesis – that it is no accident that it is exactly the most underspecified *d*-pronoun from which the complementizer is derived. I furthermore argued in detail against the claim that complementation is in fact relativization and that complementizers are in fact relative pronouns, as recently argued for by Kayne (2008, 2010). Such an assumption runs into all sorts of theoretical and empirical problems, among which Doubly Filled Comp patterns in Dutch. In the final part of this chapter, I presented a case study to support the proposal that complementizers cannot be identified with relative pronouns. On the basis of microvariation in the distribution of the elements *dat* and *die* in Dutch (long-distance) RCs, I showed that some southern Dutch varieties feature an element *die* in subject extraction (special *die*), that, although it is identical in form to the common gender distal demonstrative/relative pronoun, is best analyzed as a form variant of the complementizer, not as a relative pronoun.

CHAPTER 5

Conclusions

5.1 Summary

The main goal of this thesis was to provide a principled account of the attested *microvariation* in Dutch (long-distance) relative clauses (RCs) and related constructions, most prominently *wh*-questions (*wh*-Qs). I showed that a proper understanding of this microvariation, and in particular *doubling*, yields insight into the structure of A-bar chains (chapter 2), the syntax of RCs, the structure of the left periphery (chapter 3), and the nature of (relative) pronouns and their relation with complementizers (chapter 4). More specifically, this study has brought forth a unified analysis of doubling in long-distance A-bar dependencies, the reinstatement of the Head External Analysis of RCs, and a new implementation of an underspecification approach to so-called multipurpose pronouns.

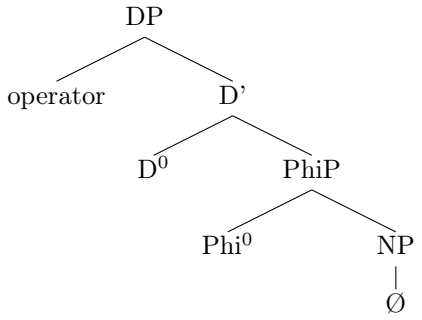
In **chapter 2**, I proposed a unified account of doubling in long-distance A-bar dependencies on the basis of new empirical data on pronominal doubling in long-distance RCs and pronominal doubling in long-distance (embedded) *wh*-Qs in Dutch. The core attested doubling data are given in (393) and (394) for a long-distance RC with a common gender human antecedent and a long-distance root *wh*-Q that questions a person respectively.

(393) de man <**die**/**wie**> ik denk <**die**/**wie**> het gedaan heeft
the man RP/who I think RP/who it done has
'the man who I think has done it'

(394) <**wie**/**wat**> denk je <**die**/**wie**> het gedaan heeft?
who/what think you RP/who it done has
'who do you think has done it?'

I argued that all long A-bar dependencies are derived by successive-cyclic movement via SpecCP of (part of) the A-bar pronoun (i.e. interrogative or relative pronoun), and that doubling comes about by spelling out more than one (part of a) copy of the A-bar pronoun at PF. More specifically, I proposed that the internal structure of A-bar pronouns includes an operator that is located in the specifier of the pronoun (cf. Szabolcsi 1994), as illustrated in (395).

(395) The structure of A-bar pronouns (= (58))



At the point in the derivation when the A-bar pronoun has reached the embedded CP domain, two possibilities emerge: either the whole pronoun (containing the operator that triggers movement) moves up, or only the operator itself moves up. The latter is possible because the pronoun and the operator in its specifier are *equally local* to the higher SpecCP. When the whole pronoun moves up, we get the movement chain as given in (396).

(396) $[_{CP}$ pronoun$_1$ ... $[_{CP}$ pronoun$_1$ ... pronoun$_1$ ...]]
full movement of the pronoun

For linearization purposes, all copies but the highest copy of the movement chain of the pronoun are deleted at PF, as illustrated in (397a). In certain cases, the copy of the pronoun in the lower SpecCP may escape this linearization requirement and be spelled out, in addition to the head of the movement chain (cf. Nunes 2004). This results in *identical doubling*, as illustrated in (397b).

(397) a. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ ~~pronoun$_1$~~ ... ~~pronoun$_1$~~ ...]]
no doubling

b. $[_{CP}$ **pronoun**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
identical doubling

When instead of the whole pronoun, only the operator moves up – i.e. *subextraction* of the operator or the lack of *pied piping* of the whole pronoun (subextraction and pied piping are two sides of the same coin) – the movement chain looks as in (398) before PF is reached.

(398) $[_{CP}$ operator$_1$ ... $[_{CP}$ pronoun$_1$ ... pronoun$_1$ ...]]
subextraction of the operator

The subextracted operator gets spelled out in the higher SpecCP. The pronoun in the lower SpecCP from which the operator subextracted needs to be spelled out as well, for reasons of *recoverability*, i.e. the full content of the pronoun cannot be recovered if it is deleted. Subextraction of the operator violates well established locality constraints on movement, like the *Condition on Extraction Domain* (CED, Huang 1982) or the *Freezing Principle* (Wexler and Culicover 1980), according to which a phrase that has undergone movement becomes opaque for extraction. I proposed that by spelling out the pronoun from which the operator subextracted, a violation of such locality principles is circumvented (cf. van Craenenbroeck and van Koppen 2008). This way of ameliorating an otherwise illicit step in the derivation, I dubbed *rescue by PF spell out* – the logical counterpart of *rescue by PF deletion* (cf. Bošković 2011 a.o.). Assuming that pronouns spell out phrases (cf. Weerman and Evers-Vermeul 2002, Barbiers et al. 2009 a.o.), spell out of the pronoun always *subsumes* spell out of the operator, as a result of which the intermediate chain link in a movement chain like the one in (398) will surface as a full pronoun. Subextraction and double spell out is abstractly illustrated in (399).

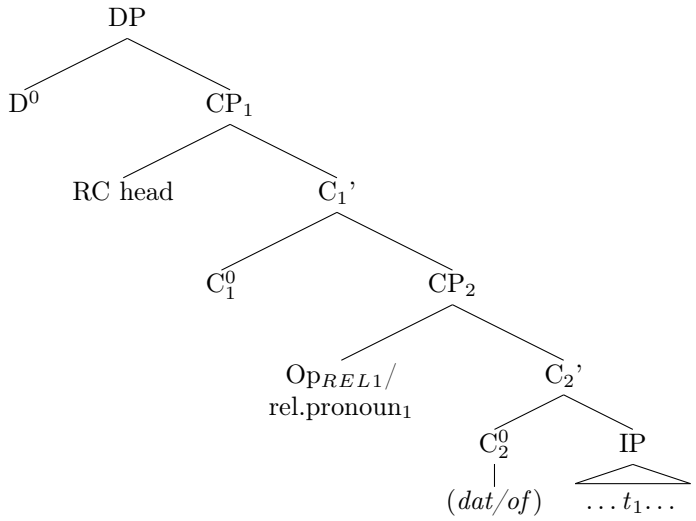
(399) $[_{CP}$ **operator**$_1$ ... $[_{CP}$ **pronoun**$_1$ ... ~~pronoun$_1$~~ ...]]
non-identical doubling

An example of this construction is the doubling pattern *wat-die/wie* in (394); the subextracted operator gets spelled out as *wat* 'what' by default – *wat* being the most underspecified A-bar pronoun in Dutch (cf. Postma 1994, Bennis 1995 a.o.). Notice that the doubling pattern *wat-die/wie* is not attested in RCs with a common gender human antecedent (cf. (393)), because such RCs cannot be introduced by the pronoun *wat*. However, this pattern seems to be attested in RCs with a neuter gender human antecedent, such as *meisje* 'girl' (e.g. *het meisje wat ik denk die het gedaan heeft* 'the girl who I think has done it'). Operator movement and double spell out thus seems not to be restricted to *wh*-Qs, but seems to be a more general strategy for forming long-distance A-bar dependencies.

The patterns of pronominal doubling in RCs and *wh*-Qs were shown to be best accounted for by the specific feature specifications of the A-bar pronouns involved (*wh*-pronouns *wie* and *wat* and *d*-pronoun *die*) and different lexicalization options and requirements. More specifically, the exact distribution of the relevant A-bar pronouns is dependent on which features they can spell out: syntactic gender (common/neuter) and/or semantic animacy (roughly human/non-human, in accordance with the *Individuation Hierarchy*, cf. Audring 2009). A late insertion model of morphology accounts for the fact that some pronouns are interchangeable in certain contexts (for some speakers), as they can be equally compatible with the structure to be lexicalized. For example, the A-bar pronouns *die* and *wie* can equally well spell out a [human] feature, as a result of which they appear to be interchangeable in the higher and lower clause of a long-distance RC with a human antecedent, and in the lower clause of a long-distance *wh*-Q that questions a person, cf. the doubling patterns in (393) and (394). The choice of which pronoun is inserted was furthermore shown to be dependent on the nature of the clause: RC or *wh*-Q. More specifically, *d*-pronouns like *die* cannot introduce *wh*-Qs due to the *wh*-requirement on the introduction of such clauses.

The most prominent analyses of RCs in current literature – so-called Head Internal Analyses (HIAs) of RC – assume that the RC head originates in the gap position inside the RC and moves to the left periphery, where it either becomes the RC head (*raising*) or gets deleted under identity with an external RC head (*matching*). **Chapter 3** started by showing that (the proposed analysis of) doubling in long-distance RCs is in fact most easily compatible with an analysis of RCs according to which the RC head does *not* originate in the gap position inside the RC: a Head External Analysis (HEA) of RCs. More specifically, under an analysis of doubling in terms of multiple copy spell out, HIAs of RCs predict that the RC head that is contained in the copy of the relative DP in the embedded SpecCP is overtly realized in doubling contexts, *quod non*. After having eliminated HIAs of RCs for doubling constructions, I proposed a specific implementation of a HEA of RCs. This analysis assumes a D-complement hypothesis (cf. Smith 1964, Kayne 1994 a.o.) – according to which the external determiner takes the relative clause CP as its complement – and a split CP domain (cf. van Craenenbroeck 2004, 2010 a.o. for Dutch). The RC head is base-generated in the highest SpecCP position of the relative clause CP itself, whereas the relative pronoun or operator targets the lower SpecCP position (cf. Schmitt 2000, Aoun and Li 2003). This is illustrated in (400). Strictly speaking, the analysis in (400) is a HIA because the RC head originates in a position inside the RC itself. However, since the RC head in the structure in (400) does not originate in the gap position inside the RC, I consider it to be a HEA.

(400) The structure of restrictive RCs in Dutch (= (226))



This analysis of Dutch RCs was shown to account for the doubling data, as well as for the full range of variation in the left periphery of Dutch RCs, particularly doubly filled COMP data.

In the remainder of chapter 3, a case was made for a HEA of RCs more generally. Even though HIAs (*raising* or *matching*) are the standard analyses of RCs in the current literature, I showed that choosing between a HEA and HIAs is not a trivial matter. The HEA is shown to fare better in many respects: case mismatches between the RC head and the relative pronoun, selectional differences between the relative pronoun and its determiner or interrogative counterpart, and locality constraint violations. Only the presence of *connectivity effects* between (material inside) the RC head and the RC internal gap seems to strongly argue against a HEA of RCs. If connectivity effects are the result of *syntactic reconstruction* – i.e. the activation of a lower copy at LF (cf. Chomsky 1993, Fox 1999b a.o.) – the presence of an (additional) RC head within the RC is needed to account for connectivity effects in RCs. However, in line with existing literature (see for instance Sharvit 1999a, Sharvit and Guerzoni 2003, Cecchetto 2005, Heycock 2005), I showed in detail that connectivity effects cannot always be accounted for by means of syntactic reconstruction. Put differently, reconstruction without movement or copies seems to be needed anyway, in RCs as well as in other configurations, such as left dislocation structures and specificational pseudoclefts. Connectivity effects are thus not a foolproof diagnostic for movement, as a result of which the presence or absence of connectivity effects in RCs provides inconclusive evidence to support any analysis of RCs.

It is well known that many languages make use of relative pronouns that are taken from another pronoun paradigm, like the demonstrative or the interrogative pronoun paradigms. Relative pronouns may thus often have more than one function and may appear in more than one syntactic configuration (*multipurpose pronouns*). The first part of **chapter 4** addressed the nature of such multipurpose pronouns – and relative pronouns in particular – and started from the null hypothesis that a multipurpose pronoun is a single lexical item that is morphosyntactically and semantically *underspecified*, as a result of which it may appear in more than one syntactic environment, and part of its meaning is determined contextually or configurationally. For demonstrative pronouns, I followed a proposal by Rooryck (2003) and argued in detail that distal demonstratives are underspecified for *location*, whereas proximal demonstratives are specified for location (as *proximal*). A consequence of this feature specification is that only distal demonstratives may function as relative pronouns. As for interrogative pronouns and the 'interrogative-relative puzzle' – i.e. the observation that many languages use relative pronouns that have the same form as interrogative pronouns (cf. Bhat 2004) – I argued that due to their underspecification for *referentiality*, *wh*-pronouns may function as relative pronouns as well.

Similarly, finite declarative complementizers are also often taken from the demonstrative or interrogative pronoun paradigms. The Dutch complementizer *dat* 'that' corresponds to the neuter gender distal demonstrative pronoun. Without going as far as claiming that complementizer *dat* and pronoun *dat* are a single lexical item in Dutch, I argued that they are diachronically related in a way that is compatible with the null hypothesis. More specifically, I showed that it is no surprise that it is exactly the most underspecified *d*-pronoun that served as a source for grammaticalization into a complementizer; *neuter* is taken to be underspecification for gender, and recall that *distal* is taken to be underspecification for location.

A recently popular view among linguists is that complementation is in fact relativization (Arsenijević 2009 a.o.) and that complementizers are in fact relative pronouns (Kayne 2008, 2010). In the second part of chapter 4, I strongly argued against this view by showing that it runs into all sorts of theoretical and empirical problems, among which its inability to properly account for the (un)attested doubly filled Comp patterns in Dutch. In addition, I provided a case study that shows the need to distinguish relative pronouns from complementizers. More specifically, some southern Dutch varieties feature the element *die* in the most deeply embedded clause of *subject* RCs: 'special *die*'. Primarily on the basis of the observation that there is a positive correlation between the distribution of special *die* and the distribution of complementizer agreement – i.e. the phenomenon by which a finite complementizer overtly agrees with the subject of the clause it introduces – I claimed that special *die* is best analyzed as an agreeing variant of the complementizer (see also Rizzi 1990 a.o.), instead of as a relative pronoun.

5.2 Empirical and theoretical contributions

This thesis builds on existing analyses of doubling, relative clauses and multipurpose pronouns. Thorough investigation of previously undiscussed data, however, has led to new empirical and theoretical insights that sometimes differ considerably from the existing claims and analyses. This section presents a short overview of the main empirical and theoretical contributions of this thesis (in order of appearance).

A comprehensive overview of the full range of variation concerning doubling in long-distance A-bar dependencies in Dutch

Whereas the SAND data (Barbiers et al. 2005, 2008a) give a good first indication of the (limits of) variation concerning doubling in long-distance A-bar dependencies, a more systematic investigation is provided by the Meertens Panel Questionnaire (MPQ) studies. The thorough examination of pronominal doubling in long-distance A-bar dependencies through the MPQs has resulted in a comprehensive overview of the full range and limits of variation concerning doubling – and concerning microvariation in (long-distance) A-bar dependencies more generally, such as variation in doubly filled Comp patterns – within the Dutch speaking language area. The large amount of new empirical data generated by the MPQ studies can serve as a good starting point for future research; not all MPQ data are discussed in this thesis. In addition, the rather new methodology that was employed in the MPQ studies – test sentences were offered to the informants in *spoken* (not written) form – could be used as a basis for future online questionnaire studies.

A unified analysis of doubling in long-distance A-bar dependencies

Much research has focused on doubling in long-distance *wh*-Qs (i.e. *wh*-copying and *wh*-scope marking), but little attention has been paid to doubling in long-distance RCs, let alone to a unified analysis of doubling in long-distance A-bar dependencies more generally. Whereas the analysis by Barbiers, Koeneman, and Lekakou (2009) provides a promising account of doubling in Dutch long-distance root *wh*-Qs, it does not extend to doubling in long-distance RCs. The very occurrence of doubling pattern *die-wie* in Dutch RCs is surprising and inconsistent with their analysis of doubling. The attestation of this doubling pattern in RCs is truly a crucial piece of data; before the MPQs established its occurrence, it was thought to be non-existent. Starting from the full range of doubling patterns in long-distance *wh*-Qs and RCs as attested in the MPQs, I propose a unified analysis of doubling in long-distance A-bar dependencies. This analysis takes doubling to be the result of multiple spell out of (a part of) the copy of the A-bar pronoun, and attributes the variation in doubling to the availability of subextraction or pied piping, and different lexicalization options for the A-bar pronoun.

In short, this thesis provides a unified analysis of two phenomena that so far have been treated independently from each other: pronoun doubling in RCs (cf. Boef 2008a, 2012b, Koopman and Sportiche 2008) and pronoun doubling in *wh*-Qs (cf. Barbiers et al. 2009, den Dikken 2009b a.o.).

PF spell out can rescue an otherwise illicit step in a derivation

A central claim in my analysis of non-identical doubling in long-distance A-bar dependencies is that a violation of the CED/Freezing Principle in terms of subextraction of the operator from a constituent in a derived position, may be ameliorated by spelling out the full constituent from which subextraction has taken place: *rescue by PF spell out*. The observation that PF *deletion* may rescue an otherwise illicit step in a derivation was first noted by Ross (1969), but the idea that PF *spell out* – the logical counterpart of PF deletion – may rescue an illicit step in a derivation is rather new (see also van Craenenbroeck and van Koppen 2008). The rescue by PF spell out approach to non-identical doubling is restricted to violations that are induced by subextraction.

Syntax can be the locus of microvariation

Within the Minimalist Program (Chomsky 1995 *et passim*), syntactic principles are assumed to be invariable. Apparent syntactic (micro)variation therefore needs to be reduced to the lexicon or vocabulary (i.e. variation in morphosyntactic features) and/or the level of PF (i.e. variation in the lexicalization of a structure). The locus of most of the microvariation that is discussed in this thesis is indeed the lexicon/vocabulary or PF. First, the variation regarding the choice for which relative pronoun is used with a particular antecedent as discussed in chapter 2, could be reduced to variation in lexicalization preferences at PF, e.g. spelling out syntactic gender versus spelling out semantic animacy. Second, the variation in the patterns of subject/object asymmetries in long-distance RCs as discussed in chapter 4, could be reduced to the lexicon/vocabulary: the availability of spelling out the agreement relation between the extracted subject and the lowest C head in RCs as *die* (i.e. the availability of *die* as a complementizer in the lexicon/vocabulary), and the availability of null relative pronouns.

As shown in detail at the end of chapter 2, part of the microvariation regarding non-identical doubling must be accounted for in syntax, namely in terms of the optionality of subextraction or pied piping; I show explicitly that the effects of subextraction or pied piping cannot be reduced to PF. The observation that some (micro)variation needs to be accounted for in terms of the size of a constituent that moves/copies in syntax is not new (cf. Barbiers 2009, Barbiers et al. 2009). In fact, my proposal fits in nicely with other proposals that attribute syntactic variation to the so-called *pied piping parameter*, cf. Koster (2000b), Koopman and Szabolcsi (2000) (see also Ross 1967, van Riemsdijk 1978).

The reinstatement of the Head External Analysis of relative clauses

A Head Internal Analysis of RCs (*raising* or *matching*) is the standard analysis of RCs in current syntactic literature. I show that a HIA faces several serious problems that have often been overlooked or that have gone previously unnoticed, like the availability of doubling of the relative pronoun in long-distance RCs in Dutch. I furthermore show that such problems are not encountered by a specific implementation of the traditional Head External Analysis of RCs.

Connectivity effects are not a foolproof diagnostic for movement

The presence of connectivity effects between (material inside) the RC head and the RC internal gap position has been the most prominent argument in favor of HIAs of RCs. An account of connectivity effects in terms of *syntactic reconstruction* (i.e. interpreting a lower copy at LF) requires the presence of a copy of the RC head inside the RC. However, I show that connectivity effects cannot always be accounted for by means of syntactic reconstruction. Put differently, connectivity effects are not a foolproof diagnostic for movement: some syntactic environments show connectivity effects without there plausibly being movement (or c-command). How exactly such connectivity effects are to be accounted for (if not by syntactic reconstruction) remains a topic for future research. Let me only point out here that several semantic accounts of connectivity have been proposed in the literature (see for instance Cresti 1995, Rullmann 1995, Sharvit 1999a, Sternefeld 2001, Ruys 2011).

The claim that connectivity effects are not a foolproof diagnostic for movement is a very strong claim that has far-reaching consequences. If true, it forces a reevaluation of all phenomena that have been argued to involve movement because they show connectivity effects. Put differently, if this claim holds true, an important diagnostic for movement in current linguistic theorizing is lost.

A multipurpose pronoun has a single underspecified lexical entry

The idea that a so-called multipurpose pronoun is *underspecified*, as a result of which it may appear in more than one syntactic environment and part of its meaning is determined contextually or configurationally, is not new (cf. Postma 1994 a.o.). I propose a specific implementation of the underspecification approach to multipurpose pronouns for relative pronouns, according to which *wh*-pronouns may act as relative pronouns because they are *underspecified* for referentiality, and distal demonstrative pronouns may act as relative pronouns because they are *underspecified* for location (cf. Rooryck 2003). Similarly, I argue that although the Dutch complementizer *dat* and the demonstrative/relative pronoun *dat* are not the same lexical item, they are diachronically related in a way that is compatible with the underspecification hypothesis: the finite declarative complementizer corresponds to the most underspecified *d*-pronoun, i.e. it is underspecified for number, gender, and location. I leave for future research the exact nature and properties of the mechanism(s) by

which the full meaning of underspecified pronouns is determined contextually or configurationally.

Relativization is pervasive but not ubiquitous in embedded clauses

Recently, it has been claimed that all complementation should be equated with relativization (cf. Kayne 2008, 2010, Arsenijević 2009). Although it has been plausibly argued that relativization (operator movement) is pervasive in certain types of embedded clauses – like temporal and conditional adverbial clauses (cf. Bhatt and Pancheva 2006, Haegeman 2009, 2010 a.o.), factive complement clauses (cf. Kiparsky and Kiparsky 1971 and many others since), and referential complement clauses (cf. de Cuba and Ürögdi 2010, Haegeman and Ürögdi 2010) – I conclude on the basis of theoretical as well as empirical arguments that the unification of complementation and relativization cannot be correct.

Special *die* is a complementizer, not a relative pronoun

In contrast to previous accounts of special *die* in Dutch – which take it to be a relative pronoun (cf. Bennis and Haegeman 1984, Koopman and Sportiche 2008) – I argue that special *die* is an (agreeing variant of the) finite declarative complementizer (cf. Rizzi 1990 a.o.). This claim is (in part) based on the attested correlation between the presence of special *die* and the availability of complementizer agreement (cf. SAND1 data: Barbiers et al. 2005).

Bibliography

Abels, Klaus. 2003. Successive Cyclicity, Anti-locality and Adposition Stranding. Doctoral Dissertation, University of Connecticut.

Abney, Steven. 1987. The English Noun Phrase in its Sentential Aspect. Doctoral Dissertation, Massachusetts Institute of Technology.

Aboh, Enoch. 2004. *The Morphosyntax of Complement-Head Sequences. Clause Structure and Word Order Patterns in Kwa*. New York: Oxford University Press.

Aboh, Enoch. 2011. Eliminating D. Handout of a talk presented at the Syntax-Interface Meeting, Universiteit Utrecht, November 2011.

Adger, David, and Gillian Catriona Ramchand. 2005. Merge and Move: Wh-Dependencies Revisited. *Linguistic Inquiry* 36:161–193.

Åfarli, Tor A. 1994. A promotion analysis of restrictive relative clauses. *The Linguistic Review* 11:81–100.

Alber, Birgit. 2008. Tyrolean A-Bar Movement: Doubling and Resumptive Pronoun Structures. In *Microvariation in Syntactic Doubling*, ed. Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, Syntax and Semantics 36, 141–170. Bingley: Emerald.

Alexiadou, Artemis, Paul Law, André Meinunger, and Chris Wilder, ed. 2000. *The Syntax of Relative Clauses*. Linguistik Aktuell/Linguistics Today 32. Amsterdam/Philadelphia: John Benjamins.

Alexopoulou, Dora. 2006. Free and Restrictive Relative Clauses in Greek. In *Proceedings of the 17th International Conference of Theoretical and Applied Linguistics, April 2005*, 85–93. Aristotle University of Thessaloniki, Greece.

Aoun, Joseph, and Yen-hui Audrey Li. 2003. *Essays on the Representational and Derivational Nature of Grammar. The Diversity of Wh-Constructions*. Linguistic Inquiry Monographs 40. Cambridge, MA: MIT Press.

Arsenijević, Boban. 2009. Clausal complementation as relativization. *Lingua* 119:39–50.

Audring, Jenny. 2006. Genusverlies en de betekenis van voornaamwoorden. In *Nederlands tussen Duits en Engels*, ed. Matthias Hüning, Arie Verhagen, Ulrike Vogl, and Ton van der Wouden, 71–88. Leiden: Stichting Neerlandistiek Leiden.

Audring, Jenny. 2008. Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology* 18:93–116.

Audring, Jenny. 2009. Reinventing Pronoun Gender. Doctoral Dissertation, Vrije Universiteit Amsterdam. LOT Dissertation Series 227.

Baker, Mark C. 1985. The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16:373–415.

Baltin, Mark R. 1984. Extraposition Rules and Discontinuous Constituents. *Linguistic Inquiry* 15:157–163.

Baltin, Mark R. 2010. The Nonreality of Doubly Filled Comps. *Linguistic Inquiry* 41:331–335.

Barbiers, Sjef. 2000. The Right Periphery in SOV languages. In *The Derivation of VO and OV*, ed. Peter Svenonius, Linguistik Aktuell/Linguistics Today 31, 181–218. Amsterdam/Philadelphia: John Benjamins.

Barbiers, Sjef. 2006. Er zijn grenzen aan wat je kunt zeggen. Inaugural lecture, Universiteit Utrecht.

Barbiers, Sjef. 2008. Microvariation in Syntactic Doubling – An Introduction. In *Microvariation in Syntactic Doubling*, ed. Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, Syntax and Semantics 36, 1–34. Bingley: Emerald.

Barbiers, Sjef. 2009. Locus and limits of syntactic microvariation. *Lingua* 119:1607–1623.

Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther de Vogelaer, and Margreet van der Ham. 2008a. *Syntactic Atlas of the Dutch Dialects, Volume 2*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, and Hans Bennis. 2007. The Syntactic Atlas of the Dutch Dialects. A discussion of choices in the SAND-project. *Nordlyd (Tromsø University Working Papers on Language and Linguistics)* 34.1:53–72.

Barbiers, Sjef, Hans Bennis, Gunther de Vogelaer, Magda Devos, and Margreet van der Ham. 2005. *Syntactic Atlas of the Dutch Dialects, Volume 1*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, Olaf Koeneman, and Marika Lekakou. 2009. Syntactic doubling and the structure of *wh*-chains. *Journal of Linguistics* 45:1–46.

Barbiers, Sjef, Olaf Koeneman, and Marika Lekakou. 2010. De linkerperiferie

gaat de dialoog aan. Een antwoord op Den Dikken. *Nederlandse Taalkunde* 15:312–314.

Barbiers, Sjef, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, ed. 2008b. *Microvariation in Syntactic Doubling*. Syntax and Semantics 36. Bingley: Emerald.

Bayer, Josef. 2004. Decomposing the left periphery. Dialectal and cross-linguistic evidence. In *The Syntax and Semantics of the Left Periphery*, ed. Horst Lohnstein and Susanne Trissler, 59–95. Berlin: Mouton de Gruyter.

Bayer, Josef, and Ellen Brandner. 2008. On Wh-Head-Movement and the Doubly-Filled-Comp Filter. In *Proceedings of WCCFL 26*, ed. Charles B. Chang and Hannah J. Haynie, 87–95. Somerville, MA: Cascadilla Proceedings Project.

Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17:1–36.

Bennis, Hans. 1983. A case of restructuring. In *Linguistics in the Netherlands 1983*, ed. Hans Bennis and W.U.S. van Lessen-Kloeke, 9–20. Dordrecht: Foris Publications.

Bennis, Hans. 1995. The meaning of structure: the *wat voor* construction revisited. In *Linguistics in the Netherlands 1995*, ed. Marcel den Dikken and Kees Hengeveld, 25–36. Amsterdam/Philadelphia: John Benjamins.

Bennis, Hans. 1997. Voegwoordvariaties. In *Taal in tijd en ruimte*, ed. Ariane van Santen and Marijke van der Wal, 353–364. SNL.

Bennis, Hans. 2000. On the interpretation of functional categories. In *Interface Strategies*, ed. Hans Bennis, Martin Everaert, and Eric Reuland, 37–53. KNAW.

Bennis, Hans. 2001. Tegengestelde krachten in taal. Inaugural lecture, UvA. Vossiuspers UvA.

Bennis, Hans, Norbert Corver, and Marcel den Dikken. 1998. Predication in Nominal Phrases. *Journal of Comparative Germanic Linguistics* 1:85–117.

Bennis, Hans, and Liliane Haegeman. 1984. On the status of agreement in relative clauses in West-Flemish. In *Sentential Complementation*, ed. Wim de Geest and Yvan Putseys, 33–53. Dordrecht/Cinnaminson: Foris Publications.

den Besten, Hans. 1983. On the Interaction of Root Transformations and Lexical Deletive Rules. In *On the Formal Syntax of the Westgermania*, ed. Werner Abraham, 47–131. Amsterdam/Philadelphia: Benjamins.

den Besten, Hans. 1985. The Ergative Hypothesis and Free Word Order in Dutch and German. In *Studies in German Grammar*, ed. Jindrich Toman, 23–62. Dordrecht: Foris Publications.

Bhat, D. N. S. 2004. *Pronouns*. Oxford: Oxford University Press.

Bhatt, Rajesh. 2002. The raising analysis of relative clauses: Evidence from adjectival modification. *Natural Language Semantics* 10:43–90.

Bhatt, Rajesh. 2005. In Search of Case Attraction. Handout of a LOT Summer School course (Leiden), March 2005: *The Syntax and Semantics of Nominal Modification*.

Bhatt, Rajesh, and Roumyana Pancheva. 2006. Conditionals. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 1, chapter 16, 638–687. Blackwell.

Bhatt, Rakesh, and James Yoon. 1992. On the Composition of COMP and Parameters of V2. In *Proceedings of WCCFL 10*, ed. Dawn Bates, 41–52.

Bianchi, Valentina. 1999. *Consequences of Antisymmetry*. Berlin: Mouton de Gruyter.

Bianchi, Valentina. 2000. The Raising Analysis of Relative Clauses: A Reply to Borsley. *Linguistic Inquiry* 31:123–140.

Bianchi, Valentina. 2002a. Headed relative clauses in generative syntax – part I. *Glot International* 6.7:197–204.

Bianchi, Valentina. 2002b. Headed relative clauses in generative syntax – part II. *Glot International* 6.8:1–13.

Biberauer, Theresa. 2008. Introduction. In *The Limits of Syntactic Variation*, 1–72. Amsterdam/Philadelphia: John Benjamins.

Boef, Eefje. 2008a. Long-distance relativization in varieties of Dutch. Master's thesis, Universiteit Utrecht.

Boef, Eefje. 2008b. Variatie in Nederlandse lange-afstandsrelatieven. *Taal en Tongval* 60:90–120.

Boef, Eefje. 2012a. Doubling in Dutch restrictive relative clauses: rethinking the Head External Analysis. In *Proceedings of ConSOLE XIX (2011, Groningen)*, ed. Enrico Boone, Kathrin Linke, and Maartje Schulpen, 125–149. Leiden: SOLE.

Boef, Eefje. 2012b. Long-distance relativization in varieties of Dutch. In *Proceedings of ConSOLE XVII (2009, Nova Gorica)*, ed. Camelia Constantinescu, Bert Le Bruyn, and Kathrin Linke, 77–100. Leiden: SOLE.

Boef, Eefje. to appear. Doubly filled Comp: a view from East-Flanders ('Doubly-filled Comp' de la vue de la Flandre-Orientale'). In *Actes du colloque international 'Morphologie, syntaxe et sémantique des subordonnants'*, ed. Colette Bodelot, Hana Gruet-Skrabalova, and François Trouilleux, Cahiers du LRL (Laboratoire de Recherche sur le Langage) 5. Maison des Sciences de l'Homme (MSH), Clermont-Ferrand.

Boef, Eefje, and Irene Franco. 2010. A comparative study of subordinate A'-dependencies. Paper presented at the 6th Grand Meeting for Scandinavian Dialect Syntax (ScanDiaSyn).

Borsley, Robert D. 1997. Relative Clauses and the Theory of Phrase Structure. *Linguistic Inquiry* 28:629–647.

Borsley, Robert D. 2001. More on the Raising Analysis of Relative Clauses. Ms., University of Essex.

Bošković, Zeljko. 1997. Pseudoclefts. *Studia Linguistica* 51:235–277.

Bošković, Zeljko. 2007. On the Locality and Motivation of Move and Agree: An Even More Minimal Theory. *Linguistic Inquiry* 38:589–644.

Bošković, Zeljko. 2011. Rescue by PF Deletion, Traces as (Non)interveners, and the *That*-Trace Effect. *Linguistic Inquiry* 42:1–44.

Bošković, Zeljko, and Jairo Nunes. 2007. The Copy Theory of Movement: A view from PF. In *The Copy Theory of Movement*, ed. Norbert Corver and Jairo Nunes, 13–74. Amsterdam/Philadelphia: John Benjamins.

Brachin, Pierre. 1973. Een in dubbel opzicht "vreemde" constructie? *De Nieuwe Taalgids* 66:59–60.

Brachin, Pierre. 1974. Een in dubbel opzicht "vreemde" constructie? *De Nieuwe Taalgids* 67:228–234.

Bresnan, Joan. 1970. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language* 6:297–321.

Brody, Michael. 2000. Mirror Theory: Syntactic Representation in Perfect Syntax. *Linguistic Inquiry* 31:29–56.

Bruening, Benjamin. 2004. Two Types of *Wh*-Scope Marking in Passamaquoddy. *Natural Language & Linguistic Theory* 22:229–305.

Bruening, Benjamin. 2006. Differences between the *Wh*-Scope Marking and *Wh*-Copy Constructions in Passamaquoddy. *Linguistic Inquiry* 37:25–49.

Caha, Pavel. 2007. The Superset Principle. Ms., University of Tromsø.

Caha, Pavel. 2009. The nanosyntax of case. Doctoral Dissertation, Tromsø University.

Cardinaletti, Anna. 1994. On the internal structure of pronominal DPs. *The Linguistic Review* 11:195–219.

Cardinaletti, Anna, and Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In *Clitics in the Languages of Europe*, ed. Henk van Riemsdijk, 145–233. Berlin: Mouton de Gruyter.

Carlson, Greg. N. 1977. Amount relatives. *Language* 58:520–542.

Ćavar, Damir, and Gisbert Fanselow. 1997. Split constituents in Germanic and Slavic. Presented at the International Conference on Pied-Piping, Friedrich-Schiller University, Jena.

Cecchetto, Carlo. 2001. Syntactic or Semantic Reconstruction? Evidence from

Pseudoclefts and Clitic Left Dislocation. In *Semantic Interfaces: Reference, Anaphora and Aspect*, ed. Carlo Cecchetto, Gennaro Chierchia, and Maria Teresa Guasti, 90–144. Stanford: CSLI Publications.

Cecchetto, Carlo. 2005. Reconstruction in relative clauses and the copy theory of traces. *Linguistic Variation Yearbook* 5:5–35.

Cheng, Lisa Lai-Shen. 1991. On the Typology of Wh-Questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Cheng, Lisa Lai-Shen. 2000. Moving Just the Feature. In *Wh-Scope Marking*, ed. Uli Lutz, Gereon Müller, and Arnim von Stechow, 77–99. Amsterdam: John Benjamins.

Cheng, Lisa Lai-Shen. 2001. Het is allemaal *wat*. Inaugural lecture, Universiteit Leiden.

Cheng, Lisa Lai-Shen, and Laura Downing. 2006. The prosody and syntax of Zulu relative clauses. *SOAS Working Papers in Linguistics* 15:51–63.

Cheng, Lisa Lai-Shen, and Johan Rooryck. 2000. Licensing *Wh*-in-situ. *Syntax* 3:1–19.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on Nominalization. In *Readings in English Transformational Grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, MA: Ginn.

Chomsky, Noam. 1977. On Wh-Movement. In *Formal Syntax*, ed. Peter Culicover, Thomas Wasow, and Adrian Akmajian, 71–132. New York: Academic Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris Publications. Reprint, 7th Edition. Berlin–New York: Walter de Gruyter & Co., 1993.

Chomsky, Noam. 1993. A Minimalist Program for Linguistic Theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam, and Howard Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8:425–504.

Cinque, Guglielmo. 1979. *Studi di sintassi e pragmatica*. Padova: CLESP.

Cinque, Guglielmo. 1990. *Types of A'-Dependencies*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2008. More on the indefinite character of the head of restrictive relatives. In *Selected Proceedings of the 34th Incontro di Grammatica Generativa (Special issue of the Rivista di Grammatica Generativa)*, ed. Paola Benincà, Federico Damonte, and Nicoletta Penello, volume 33, 3–24. Padova: Unipress.

Cinque, Guglielmo. 2010. Toward a unified theory of relative clauses. Handout of a talk presented at Mini-Workshop on Modifiers and Carthography. CASTL, University of Tromsø.

Citko, Barbara. 2001. *Deletion Under Identity* in Relative Clauses. In *Proceedings of NELS 31*, ed. Minjoo Kim and Uri Strauss, 131–145. GLSA, University of Massachusetts Amherst.

Cole, Peter. 1987. The Structure of Internally Headed Relative Clauses. *Natural Language & Linguistic Theory* 5:277–302.

Corbett, Greville. 1979. The agreement hierarchy. *Journal of Linguistics* 15:203–224.

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press.

Cornips, Leonie, and Willy Jongenburger. 2001a. Elicitation techniques in a Dutch syntactic dialect atlas project. In *Linguistics in the Netherlands 2001*, ed. Hans Broekhuis and Ton van der Wouden, volume 18, 57–69. Amsterdam/Philadelphia: John Benjamins.

Cornips, Leonie, and Willy Jongenburger. 2001b. Het design en de methodologie van het SAND-project. *Nederlandse Taalkunde* 215–232.

Cornips, Leonie, and Cecilia Poletto. 2005. On standardising syntactic elicitation techniques (part 1). *Lingua* 115:939–957.

Corver, Norbert. 1990. The Syntax of Left Branch Extractions. Doctoral Dissertation, Tilburg University.

Corver, Norbert. 1991. Wat voor construction is de 'wat voor'-constructie? *Spektator* 20:151–185.

Corver, Norbert. 2003. On three types of movement within the Dutch nominal domain. In *From NP to DP. Volume 1: The syntax and semantics of noun phrases*, ed. Martine Coene and Yves D'hulst, chapter 12, 297–328. Amsterdam/Philadelphia: John Benjamins.

Corver, Norbert. 2006. Freezing Effects. In *The Blackwell Companion to Syn-*

tax, ed. Martin Everaert and Henk van Riemsdijk, volume 2, chapter 28, 383–406. Oxford: Blackwell.

Corver, Norbert, and Dennis Delfitto. 1999. On the nature of pronoun movement. In *Clitics in the Languages of Europe*, ed. Henk van Riemsdijk, 799–861. Berlin: Mouton de Gruyter.

Corver, Norbert, and Marjo van Koppen. 2008. De wie? De die! Handout of a talk presented at the TIN-dag, February 2008.

van Craenenbroeck, Jeroen. 2004. Ellipsis in Dutch Dialects. Doctoral Dissertation, Universiteit Leiden. LOT Dissertation Series 96.

van Craenenbroeck, Jeroen. 2010. Complex *Wh*-Phrases Don't Move: On the Interaction between the Split CP Hypothesis and the Syntax of *Wh*-movement. In *The Complementizer Phase: Subjects and Operators*, ed. Phoevos Panagiotidis, 236–260. Oxford: Oxford University Press.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2002a. Pronominal Doubling and the Structure of the Left Periphery in Southern Dutch. In *Syntactic Microvariation*, ed. Sjef Barbiers, Leonie Cornips, and Suzanne van der Kleij, 280–304. Meertens Institute Electronic Publications in Linguistics (MIEPiL). Available at `http://www.meertens.knaw.nl/books/synmic/`.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2002b. Subject doubling in Dutch dialects. In *Proceedings of Console IX (2000, Lund)*, ed. Marjo van Koppen, Erica Thrift, Erik Jan van der Torre, and Malte Zimmermann, 54–67. Leiden: SOLE.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2007. Theorie, empirie en subjectverdubbeling: een antwoord op Haegeman (2005). *Taal en Tongval* 59:149–171.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2008. Pronominal Doubling in Dutch Dialects: Big DPs and Coordinations. In *Microvariation in Syntactic Doubling*, ed. Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, Syntax and Semantics 36, 207–249. Bingley: Emerald.

Cresti, Diana. 1995. Extraction and reconstruction. *Natural Language Semantics* 3:79–122.

de Cuba, Carlos, and Barbara Ürögdi. 2010. Clearing up the 'Facts' on Complementation. In *Proceedings of the 33rd Annual Penn Linguistics Colloquium*, 41–50. University of Pennsylvania Working Papers in Linguistics 16.1.

Dayal, Veneeta. 1994. Scope marking as indirect *wh*-dependency. *Natural Language Semantics* 2:137–170.

Dayal, Veneeta. 2000. Scope Marking: Cross-Linguistic Variation in Indirect Dependency. In *Wh-Scope Marking*, ed. Uli Lutz, Gereon Müller, and Arnim von Stechow, 157–194. Amsterdam: John Benjamins.

Déchaine, Rose-Marie, and Martina Wiltschko. 2002. Decomposing Pronouns. *Linguistic Inquiry* 33:409–442.

Dekkers, Joost. 1999. Derivations & Evaluations. On the Syntax of Subjects and Complementizers. Doctoral Dissertation, Universiteit van Amsterdam, The Hague: Holland Academic Graphics.

den Dikken, Marcel. 1995. *Particles. On the Syntax of Verb-Particle, Triadic, and Causative Constructions*. Oxford Studies in Comparative Syntax. Oxford/New York: Oxford University Press.

den Dikken, Marcel. 2003. On the morphosyntax of *wh*-movement. In *Multiple Wh-Fronting*, ed. Cedric Boeckx and Kleanthes K. Grohmann, 77–98. Amsterdam/Philadelphia: John Benjamins.

den Dikken, Marcel. 2006a. *Relators and Linkers. The Syntax of Predication, Predicate Inversion, and Copulas*. Cambridge, MA: MIT Press.

den Dikken, Marcel. 2006b. Specificational Copular Sentences and Pseudoclefts. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 4, chapter 61, 292–409. Oxford: Blackwell.

den Dikken, Marcel. 2007. Questionnaire study on Dutch *that*-trace effects: Stimuli and results. Ms., CUNY Graduate Center.

den Dikken, Marcel. 2009a. Arguments for successive-cyclic movement through SpecCP: A critical review. *Linguistic Variation Yearbook* 9:89–126.

den Dikken, Marcel. 2009b. On the nature and distribution of successive cyclicity. Adjunction, resumption, and scope marking as the roads to success in long-distance relation building. Ms., CUNY Graduate Center.

den Dikken, Marcel. 2009c. Predication and specification in the syntax of cleft sentences. Ms., CUNY Graduate Center.

den Dikken, Marcel. 2010. De linkerperiferie antwoordt. Een reactie op het artikel van Barbiers, Koeneman & Lekakou. *Nederlandse Taalkunde* 15:308–311.

den Dikken, Marcel, and Hans Bennis. 2009. Wat dacht je wat? On the strategies for the formation of long-distance dependencies in constituent questions. Public lecture at the Royal Netherlands Academy of Arts and Sciences, Amsterdam.

den Dikken, Marcel, André Meinunger, and Chris Wilder. 2000. Pseudoclefts and Ellipsis. *Studia Linguistica* 54:41–89.

Dixon, Robert M.W. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language* 27:61–112.

Erteschik-Shir, Nomi. 1973. On the nature of island constraints. Doctoral Dissertation, Massachusetts Institute of Technology.

Fabb, Nigel. 1990. The Difference between English Restrictive and Nonrestric-

tive Relative Clauses. *Journal of Linguistics* 26:57–78.

Fábregas, Antonio. 2007. The Exhaustive Lexicalization Principle. *Nordlyd (Tromsø Working Papers on Language and Linguistics) 34.2: Special issue on Space, Motion, and Result,* 165–199.

Fanselow, Gisbert. 2006. Partial *Wh*-Movement. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 3, chapter 47, 437–492. Blackwell.

Fanselow, Gisbert, and Anoop Mahajan. 2000. Towards a Minimalist Theory of Wh-Expletives, Wh-Copying, and Successive Cyclicity. In *Wh-Scope Marking*, ed. Uli Lutz, Gereon Müller, and Arnim von Stechow, 195–230. Amsterdam: John Benjamins.

Felser, Claudia. 2001. *Wh*-Expletives and Secondary Predication: German Partial *Wh*-Movement Reconsidered. *Journal of Germanic Linguistics* 13:5–38.

Felser, Claudia. 2004. *Wh*-Copying, Phases and Successive-Cyclicity. *Lingua* 114:543–574.

Fiengo, Robert, and Robert May. 1994. *Indices and Identity*. Linguistic Inquiry Monographs 24. Cambridge, MA: MIT Press.

Fischer, Silke. 2002. Reanalyzing reconstruction effects. An optimality-theoretic account of the relation between pronouns and R-expressions. In *Proceedings of Console IX (2000, Lund)*, ed. Marjo van Koppen, Erica Thrift, Erik Jan van der Torre, and Malte Zimmermann, 68–81. Leiden: SOLE.

Fox, Danny. 1999a. *Economy and Semantic Interpretation*. Linguistic Inquiry Monographs 35. Cambridge, MA: MIT Press.

Fox, Danny. 1999b. Reconstruction, Variable Binding and the Interpretation of Chains. *Linguistic Inquiry* 30:157–196.

Frampton, John, and Sam Gutmann. 2000. Agreement is Feature Sharing. Ms., Northeastern University, Boston.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100:25–50.

Freidin, Robert. 1986. Fundamental issues in the theory of binding. In *Studies in the Acquisition of Anaphora, Volume 1*, ed. Barbara Lust, 151–188. Dordrecht: D. Reidel Publishing Company.

Gallego, Ángel J. 2010. Subextraction from Phase Edges. In *The Complementizer Phase: Subjects and Operators*, ed. Phoevos Panagiotidis, 51–75. Oxford: Oxford University Press.

Gallego, Ángel J., and Juan Uriagereka. 2007. Conditions on sub-extraction. In *Coreference, Modality and Focus: Studies on the Syntax-Semantics Interface*, ed. Luis Eguren and Olga Fernández Soriano, 45–70. Amsterdam: John Benjamins.

Grohmann, Kleanthes K. 2003. *Prolific Domains: On the Anti-Locality of Movement Dependencies*. Linguistik Aktuell/Linguistics Today 66. Amsterdam: John Benjamins.

Groos, Anneke, and Henk van Riemsdijk. 1981. Matching effects with free relatives: A parameter of core grammar. In *Theory of markedness in generative grammar: Proceedings of the 1979 GLOW conference*, ed. Adriana Belletti, Luciana Brandi, and Luigi Rizzi, 171–216. Pisa: Scuola Normale Superiore.

Grosu, Alexander, and Fred Landman. 1998. Strange relatives of the third kind. *Natural Language Semantics* 6:125–170.

Guilliot, Nicolas, and Nouman Malkawi. 2007. Reconstruction without movement. In *Coreference, Modality and Focus: Studies on the Syntax-Semantics Interface*, ed. Luis Eguren and Olga Fernández Soriano, 113–132. Amsterdam: John Benjamins.

Hackl, Martin, and Jorie Koster-Moeller. 2008. D-raising in restrictive relative clauses. XVIII Colloquium on Generative Grammar.

Haegeman, Liliane. 1983. *Die* and *dat* in West-Flemish relative clauses. In *Linguistics in the Netherlands 1983*, ed. Hans Bennis and W.U.S. van Lessen-Kloeke, 83–91. Dordrecht: Foris Publications.

Haegeman, Liliane. 1992. *Theory and Description in Generative Syntax. A Case Study in West Flemish*. Cambridge, MA: Cambridge University Press.

Haegeman, Liliane. 2004a. DP periphery and clausal periphery: possessor doubling in West-Flemish? In *Peripheries*, ed. David Adger, Cécile de Cat, and George Tsoulas, 211–240. Dordrecht: Kluwer Academic Publishers.

Haegeman, Liliane. 2004b. Verdubbeling van subjectpronomina in de Zuid-Nederlandse dialecten: een reactie uit Lapscheure. *Taal en Tongval* 56:119–159.

Haegeman, Liliane. 2009. The movement analysis of temporal adverbial clauses. *English Language and Linguistics* 13:385–408.

Haegeman, Liliane. 2010. The internal syntax of adverbial clauses. *Lingua* 120:628–648.

Haegeman, Liliane, and Barbara Ürögdi. 2010. Referential CPs and DPs: An operator movement account. *Theoretical Linguistics* 36:111–152.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten C. van den Doorn. 1997. *Algemene Nederlandse Spraakkunst (ANS)*. Groningen: Wolters-Noordhoff.

Halle, Morris. 1997. Distributed Morphology: Impoverishment and Fission. In *MIT Working Papers in Linguistics 30: Papers at the Interface*, ed. Benjamin Bruening, Yoonjung Kang, and Martha McGinnis, 425–449.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In *The View from Building 20: Essays in Linguistics in Honor*

of Sylvain Bromberger, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, MA: MIT Press.

Halle, Morris, and Alec Marantz. 1994. Some Key Features of Distributed Morphology. *MIT Working Papers in Linguistics 21: Papers on Phonology and Morphology,* 275–288.

Harley, Heidi, and Rolf Noyer. 1999. Distributed Morphology. *Glot International* 4:3–9.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature geometric analysis. *Language* 78:482–526.

Hartman, Jeremy. 2009. Intervention in Tough-Constructions. In *Proceedings of NELS 39*.

Haslinger, Irene. 2007. The Syntactic Location of Events: Aspects of Verbal Complementation in Dutch. Doctoral Dissertation, Universiteit van Tilburg. LOT Dissertation Series 169.

Heck, Fabian. 2005. Gegen Kopfanhebung in deutschen Relativsätzen. Handout GGS Tübingen.

Heck, Fabian, and Juan Cuartero. 2008. Long Distance Agreement in Relative Clauses. In *Varieties of Competition, Linguistische Arbeitsberichte (Uni Leipzig) 87*, ed. Fabian Heck, Gereon Müller, and Jochen Trommer, 13–48. University of Leipzig.

Heim, Irene. 1982. The Semantics of Definite and Indefinite Noun Phrases. Doctoral Dissertation, University of Massachusetts Amherst.

Heim, Irene. 1987. Where Does the Definiteness Restriction Apply? Evidence from the Definiteness of Variables. In *The Representation of (In)definiteness*, ed. Eric Reuland and Alice ter Meulen, chapter 2, 21–42. Cambridge, MA: MIT Press.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell Textbooks in Linguistics. Oxford: Blackwell.

Henderson, Brent. 2007. Matching and Raising Unified. *Lingua* 117:202–220.

Heycock, Caroline. 1995. Asymmetries in Reconstruction. *Linguistic Inquiry* 26:547–570.

Heycock, Caroline. 2005. On the interaction of adjectival modifiers and relative clauses. *Natural Language Semantics* 13:359–382.

Heycock, Caroline. 2011. Relative Reconstructions. Handout of a talk presented at the workshop 'Reconstruction Effects in Relative Clauses', Zentrum für Allgemeine Sprachwissenschaft (ZAS), Berlin, July 2011.

Hicks, Glynn. 2009. *Tough*-Constructions and Their Derivation. *Linguistic Inquiry* 40:535–566.

Hiemstra, Inge. 1986. Some aspects of *wh*-questions in Frisian. *North-Western*

European Language Evolution (NOWELE) 8:97–110.

Higginbotham, James. 1993. Interrogatives. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 195–227. Cambridge, MA: MIT Press.

Higginbotham, James, and Robert May. 1981. Questions, quantifiers and crossing. *The Linguistic Review* 1:41–80.

Higgins, Francis Roger. 1979. *The pseudo-cleft construction in English*. New York: Garland Publishing.

Hladnik, Marko. 2010. Restrictive relative clauses in Slovene and its spoken varieties. Master's thesis, Universiteit Utrecht.

Hoeksema, Jack. 1986. An Account of Relative Clauses with Split Antecedents. In *Proceedings of WCCFL 5*, ed. Mary Dalrymple, Jeffrey Goldberg, Kristin Hanson, Michael Inman, Christopher Piñón, and Stephen Wechsler, 68–86.

Hoekstra, Eric. 1993a. Dialectal variation inside CP as parametric variation. In *Dialektsyntax*, ed. Werner Abraham and Josef Bayer, 161–179. Opladen: Westdeutscher Verlag.

Hoekstra, Eric. 1993b. On the Parametrisation of Functional Projections in CP. In *Proceedings of NELS 23*, ed. Amy J. Schafer, 191–204. GLSA, University of Massachusetts Amherst.

Hoekstra, Eric. 1994. Overtollige voegwoorden en de volgorde *of + interrogativum/relativum*. *De Nieuwe Taalgids* 87:314–321.

Hoekstra, Eric. 1999. On D-pronouns and the Movement of Topic Features. *Folia Linguistica* 33:59–74.

Hoekstra, Eric, and Jan-Wouter Zwart. 1994. De Structuur van CP. Functionele Projecties voor Topics en Vraagwoorden in het Nederlands. *Spektator* 23:191–212.

Hoekstra, Jarich, and László Marácz. 1989. Some implications of I-to-C movement in Frisian. In *Linguistics in the Netherlands 1989*, ed. Hans Bennis and Ans van Kemenade, 81–99. Dordrecht: Foris Publications.

Hoekstra, Teun. 2004. ECP, tense and islands. In *Aguments and Structure. Studies on the Architecture of the Sentence*, ed. Rint Sybesma, Sjef Barbiers, Marcel den Dikken, Jenny Doetjes, Gertjan Postma, and Guido Vanden Wyngaerd, Studies in Generative Grammar 67, 191–215. Berlin: Mouton de Gruyter.

Horn, George. 2003. Idioms, metaphors and syntactic mobility. *Journal of Linguistics* 39:245–273.

van der Horst, Joop, and Kees van der Horst. 1999. *Geschiedenis van het Nederlands in de twinstigste eeuw*. Den Haag: Sdu.

van der Horst, Joop M. 2008. *Geschiedenis van de Nederlandse Syntaxis*. Leu-

ven: Universitaire Pers Leuven.

Huang, C.-T. James. 1982. Logical Relations in Chinese and the Theory of Grammar. Doctoral Dissertation, Massachusetts Institute of Technology.

Hulsey, Sarah, and Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics* 14:111–137.

Jackendoff, Ray S. 1977. *X-Bar Syntax: A Study of Phrase Structure*. Linguistic Inquiry Monographs 2. Cambridge, MA: MIT Press.

Jackendoff, Ray S. 1997. *The Architecture of the Language Faculty*. Linguistic Inquiry Monographs 28. Cambridge, MA: MIT Press.

Jacobson, Pauline. 1994. Binding Connectivity in Copular Sentences. In *Proceedings of SALT IV*, ed. Mandy Harvey and Lynn Santelmann, 161–178. Ithaca, NY: Cornell University, CLC Publications.

Jayaseelan, K.A. 2008. Question particles and disjunction. Ms., The English and Foreign Languages University, Hyderabad.

Jónsson, Jóhannes Gísli. 2008. Preposition Reduplication in Icelandic. In *Microvariation in Syntactic Doubling*, ed. Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, Syntax and Semantics 36, 403–417. Bingley: Emerald.

van Kampen, Jacqueline. 1997. *First Steps in Wh-movement*. Delft: Eburon.

van Kampen, Jacqueline. 2007. Relative Agreement in Dutch. In *Linguistics in the Netherlands 2007*, ed. Bettelou Los and Marjo van Koppen, 112–125. Amsterdam/Philadelphia: John Benjamins.

van Kampen, Jacqueline. 2010. The learnability of A-bar chains. In *The Linguistics Enterprise. From knowledge of language to knowledge in linguistics*, ed. Martin Everaert, Tom Lentz, Hannah de Mulder, Øystein Nilsen, and Arjen Zondervan, Linguistik Aktuell/Linguistics Today 150, 115–140. Amsterdam/Philadelphia: John Benjamins.

Kayne, Richard S. 1976. French relative "que". In *Current studies in Romance linguistics*, ed. Marta Luján and Fritz Hensey, 255–299. Washington, D.C.: Georgetown University Press.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Linguistic Inquiry Monographs 25. Cambridge, MA: MIT Press.

Kayne, Richard S. 2000. *Parameters and Universals*. New York: Oxford University Press.

Kayne, Richard S. 2005. Some Notes on Comparative Syntax, with Special Reference to English and French. In *The Oxford Handbook of Comparative Syntax*, ed. Guglielmo Cinque and Richard S. Kayne, 3–69. New York: Oxford University Press.

Kayne, Richard S. 2008. Antisymmetry and the Lexicon. *Linguistic Variation*

Yearbook 8:1–31.

Kayne, Richard S. 2010. Why Isn't *This* a Complementizer. In *Comparisons and Contrasts*, chapter 10, 190–227. Oxford: Oxford University Press.

Kayne, Richard S. 2011. More on relative pronouns. Handout of a talk presented at the Syntax Circle, Universiteit Leiden, October 2011.

Kennedy, Christopher. 2000. Comparative (Sub)deletion and Ranked, Violable Constraints in Syntax. In *Proceedings of NELS 30*, ed. Masako Hirotani, Andries Coetzee, Nancy Hall, and Ji-yung Kim, 389–413. GLSA, University of Massachusetts Amherst.

Kester, Ellen-Petra. 1996. The Nature of Adjectival Inflection. Doctoral Dissertation, Universiteit Utrecht.

Kiparsky, Paul. 1973. 'Elsewhere' in Phonology. In *A Festschrift for Morris Halle*, ed. Stephen R. Anderson and Paul Kiparsky, 93–106. New York: Holt, Rinehart & Winston.

Kiparsky, Paul. 1995. Indo-European Origins of Germanic Syntax. In *Clause Structure and Language Change*, ed. Ian Roberts and Adrian Battye, 140–169. Oxford: Oxford University Press.

Kiparsky, Paul, and Carol Kiparsky. 1971. Fact. In *Semantics. An Interdisciplinary Reader in Philosophy Linguistics and Psychology*, ed. Danny D. Steinberg and Leon A. Jakobovits, 345–369. Cambridge: Cambridge University Press.

Klima, E.S. 1964. Studies in Diachronic Transformational Syntax. Doctoral Dissertation, Harvard University.

Koopman, Hilda. 1996. The Spec-Head Configuration. In *Syntax at Sunset: UCLA Working Papers in Linguistics, 1996*, ed. Edward Garrett and Felicia Lee, 37–64. Los Angeles: UCLA.

Koopman, Hilda. 1999. The internal and external distribution of pronominal DPs. In *Beyond Principles and Parameters: Essays in Memory of Osvaldo Jaeggli*, ed. Kyle Johnson and Ian Roberts, 91–132. Dordrecht: Kluwer.

Koopman, Hilda. 2000. *The Syntax of Specifiers and Heads*. London: Routledge.

Koopman, Hilda, and Dominique Sportiche. 2008. The *que/qui* alternation: new analytical directions. (to appear in *Functional structure from top to toe: A Festschrift for Tarald Taraldsen*, ed. Peter Svenonius. Oxford: Oxford University Press).

Koopman, Hilda, and Anna Szabolcsi. 2000. *Verbal Complexes*. Cambridge, MA: MIT Press.

van Koppen, Marjo. 2005. One probe – two goals: aspects of agreement in Dutch dialects. Doctoral Dissertation, Universiteit Leiden. LOT Dissertation Series 105.

van Koppen, Marjo. 2007. Agreement with (the internal structure of) copies of movement. In *The Copy Theory of Movement*, ed. Norbert Corver and Jairo Nunes, 327–350. Amsterdam/Philadelphia: John Benjamins.

Kornfilt, Jaklin. 2009. Subject-Agreement correlations and their syntactic effects in some Turkic relative clauses. *Turkic Languages* 13:70–96.

Koster, Jan. 2000a. Extraposition as Parallel Construal. Ms., Rijksuniversiteit Groningen.

Koster, Jan. 2000b. Pied Piping and the Word Orders of English and Dutch. In *Proceedings of NELS 30*, ed. Masako Hirotani, Andries Coetzee, Nancy Hall, and Ji-yung Kim, 415–426. GLSA, University of Massachusetts Amherst.

Koster, Jan. 2009. IM not Perfect: The Case against Copying. Ms., Rijksuniversiteit Groningen.

Koster-Moeller, Jorie. 2012. Internal DP heads in restrictive relative clauses. In *Proceedings of ConSOLE XVII (2009, Nova Gorica)*, ed. Camelia Constantinescu, Bert Le Bruyn, and Kathrin Linke, 209–230. Leiden: SOLE.

Krapova, Iliyana. 2010. Bulgarian relative and factive clauses with an invariant complementizer. *Lingua* 120:1240–1272.

Larson, Richard K. 1985. Bare-NP Adverbs. *Linguistic Inquiry* 16:595–621.

Larson, Richard K. 1987. "Missing Prepositions" and the Analysis of English Free Relative Clauses. *Linguistic Inquiry* 18:239–266.

Lasnik, Howard. 1998. Some Reconstruction Riddles. In *University of Pennsylvania Working Papers in Linguistics*, volume 5, 83–98.

Lasnik, Howard. 2012. A Reconsideration of Island Repair by Deletion. Handout of a talk presented at Leiden University, April 2012.

Lasnik, Howard, and Robert Fiengo. 1974. Complement Object Deletion. *Linguistic Inquiry* 5:535–572.

Lebeaux, David. 1988. Language Acquisition and the Form of Grammar. Doctoral Dissertation, University of Massachusetts Amherst.

Lebeaux, David. 1990. Relative clauses, licensing, and the nature of derivation. In *Proceedings of NELS 20*, 318–332. GLSA, University of Massachusetts Amherst.

Lees, Robert B. 1960. *The Grammar of English Nominalizations*. The Hague: Mouton.

Lees, Robert B. 1961. The Constituent Structure of Noun Phrases. *American Speech* 36:159–168.

Lehmann, Christian. 1984. *Der relativsatz: Typologie seiner strukturen; theorie seiner funktionen; kompendium seiner grammatik*. Language Universals Series 3. Tübingen: Gunter Narr Verlag.

Leu, Thomas. 2008. The Internal Syntax of Determiners. Doctoral Dissertation,

New York University, Published as Groninger Arbeiten zur Germanistischen Linguistik 47.

Leu, Thomas. 2010. Generalized x-to-C in Germanic. Ms., Yale University.

Linebarger, Marcia. 1980. The grammar of negative polarity. Doctoral Dissertation, Massachusetts Institute of Technology.

Longobardi, Giuseppe. 1994. Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form. *Linguistic Inquiry* 25:609–665.

Lutz, Uli, Gereon Müller, and Arnim von Stechow, ed. 2000. *Wh-Scope Marking*. Amsterdam: John Benjamins.

Manzini, Maria Rita. 2010. The Structure and Interpretation of (Romance) Complementizers. In *The Complementizer Phase: Subjects and Operators*, ed. Phoevos Panagiotidis, 167–199. Oxford: Oxford University Press.

Manzini, Maria Rita, and Leonardo M. Savoia. 2003. The nature of complementizers. *Rivista di Grammatica Generativa* 28:87–110.

Marantz, Alec. 1982. Re Reduplication. *Linguistic Inquiry* 13:435–482.

Mayr, Clemens. 2010. On the Necessity of Phi-features: The Case of Bavarian Subject Extraction. In *The Complementizer Phase: Subjects and Operators*, ed. Phoevos Panagiotidis, 117–142. Oxford: Oxford University Press.

McCloskey, James. 1990. Resumptive Pronouns, A'-Binding and Levels of Representation in Irish. In *The Syntax of the Modern Celtic Languages. Syntax and Semantics Volume 23*, ed. Randall Hendrick, 199–248. Academic Press.

McCloskey, James. 2006. Resumption. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 4, chapter 55, 94–117. Oxford: Blackwell.

McDaniel, Dana. 1989. Partial and Multiple *Wh*-Movement. *Natural Language & Linguistic Theory* 7:565–604.

Munn, Alan. 1994. A Minimalist account of reconstruction asymmetries. In *Proceedings of NELS 24*, ed. Merce Gonzalez, 397–410. GLSA, University of Massachusetts Amherst.

Neeleman, Ad, and Hans van de Koot. 2006. Syntactic Haplology. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 4, chapter 69, 684–710. Oxford: Blackwell.

Neeleman, Ad, and Hans van de Koot. 2010. A Local Encoding of Syntactic Dependencies and Its Consequences for the Theory of Movement. *Syntax* 13:331–372.

Neeleman, Ad, and Krista Szendröi. 2007. Radical Pro Drop and the Morphology of Pronouns. *Linguistic Inquiry* 38:671–714.

Noguchi, Tohru. 1997. Two types of pronouns and variable binding. *Language* 73:770–797.

Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language* 70:491–538.

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Linguistic Inquiry Monographs 43. Cambridge, MA: MIT Press.

Ott, Dennis. 2011. An ellipsis approach to Contrastive Left-dislocation. Ms., Rijksuniversiteit Groningen.

Pankau, Andreas. 2009. Wh-copying in German and its theoretical implications. Handout of a talk presented at the Workshop on Long-Distance Dependencies, Meertens Instituut, December 2009.

Perlmutter, David. 1971. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart & Winston.

Pesetsky, David. 1987. Wh-in-Situ: Movement and Unselective Binding. In *The Representation of (In)definiteness*, ed. Eric Reuland and Alice ter Meulen, 98–129. Cambridge, MA: MIT Press.

Pesetsky, David. 1998. Some Optimality Principles of Sentence Pronunciation. In *Is the Best Good Enough? Optimality and Competition in Syntax*, ed. Pilar Barbosa, Danny Fox, Martha McGinnis, and David Pesetsky, 337–383. Cambridge, MA: MIT Press.

Pesetsky, David. 2000. *Phrasal Movement and its Kin*. Linguistic Inquiry Monographs 37. Cambridge, MA: MIT Press.

Pesetsky, David, and Esther Torrego. 2007. The Syntax of Valuation and the Interpretability of Features. In *Phrasal and Clausal Architecture: Syntactic derivation and interpretation*, ed. Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 262–294. Amsterdam: John Benjamins.

Platzack, Christer. 2000. A Complement-of-N° Account of Restrictive and Non-Restrictive Relatives: The Case of Swedish. In *The Syntax of Relative Clauses*, ed. Artemis Alexiadou, Paul Law, André Meinunger, and Chris Wilder, Linguistik Aktuell/Linguistics Today 32, 265–308. Amsterdam: John Benjamins.

Poletto, Cecilia. 2006. Doubling as a Spare Movement Strategy. Paper presented at the Workshop on Doubling, Meertens Instituut, March 2006.

Poletto, Cecilia, and Jean-Yves Pollock. 2004. On wh-clitics and wh-doubling in French and some North Eastern Italian dialects. *Probus* 16:241–272.

Pollard, Carl J., and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Postal, Paul. 1969. On So-Called "Pronouns" in English. In *Modern Studies in English: Readings in Transformational Grammar*, ed. David A. Reibel and Sanford A. Schane, 201–224. Englewood Cliffs, NJ: Prentice-Hall.

Postma, Gertjan. 1994. The indefinite reading of WH. In *Linguistics in the Netherlands 1994*, ed. Reineke Bok-Bennema and Crit Cremers, 187–198.

Amsterdam/Philadelphia: John Benjamins.

Postma, Gertjan. 1997. The nature of complementizer *az* in Eastern Yiddish – Implications for its position within Germanic. Paper presented at the Workshop on Yiddish, Paris, June 1997.

Postma, Gertjan. 2006. Van *groter dan* naar *groter als* – structurele oorzaken voor het verval van het comparatieve voegwoord *dan*. *Nederlandse Taalkunde* 11:2–22.

Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Rackowski, Andrea, and Norvin Richards. 2005. Phase Edge and Extraction: A Tagalog Case Study. *Linguistic Inquiry* 36:565–599.

Ramchand, Gillian Catriona. 2008a. Lexical Items in Complex Predications: Selection as Underassociation. *Nordlyd (Tromsø Working Papers on Language and Linguistics) 35: Special issue on Complex Predication,* 115–141.

Ramchand, Gillian Catriona. 2008b. *Verb Meaning and the Lexicon. A First-Phase Syntax*. Cambridge Studies in Linguistics 116. Cambridge: Cambridge University Press.

Reinhart, Tanya. 1980. On the Position of Extraposed Clauses. *Linguistic Inquiry* 11:621–624.

Reinhart, Tanya. 1981. A second Comp position. In *Theory of markedness in generative grammar. Proceedings of the 1979 GLOW conference*, ed. Adriana Belletti, Luciana Brandi, and Luigi Rizzi, 517–557. Scuola Normale Superiore.

Rett, Jessica. 2006. Pronominal vs. determiner *wh*-words: evidence from the copy construction. In *Empirical Issues in Syntax and Semantics 6*, ed. Olivier Bonami and Patricia Cabredo Hofherr, 355–374. Paris: Colloque de Syntaxe et Sémantique à Paris. Available at `http://www.cssp.cnrs.fr/eiss6/`.

van Riemsdijk, Henk. 1978. A case study in syntactic markedness: The binding nature of prepositional phrases. Doctoral Dissertation, Universiteit van Amsterdam.

van Riemsdijk, Henk. 1983. Correspondence Effects and the Empty Category Principle. In *Studies in Generative Grammar and Language Acquisition: A Report on Recent Trends in Linguistics*, ed. Yukio Otsu, Henk van Riemsdijk, Kazuko Inoue, Akio Kamio, and Noriko Kawasaki, 5–16. Tokyo: International Christian University.

van Riemsdijk, Henk. 2006. Free Relatives. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 2, chapter 27, 338–382. Oxford: Blackwell.

Rigterink, Mirjam. 2005. About the properties of demonstratives in evaluative constructions. Handout of a talk presented at the workshop Nominal

Anaphors, Universiteit Utrecht.

Rijkhoek, Paulien Dea. 1998. On Degree Phrases and Result Clauses. Doctoral Dissertation, Rijksuniversiteit Groningen. Groningen Dissertations in Linguistics 27.

Ritter, Elizabeth. 1995. On the Syntactic Category of Pronouns and Agreement. *Natural Language & Linguistic Theory* 13:405–443.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris Publications.

Rizzi, Luigi. 1990. *Relativized Minimality*. Linguistic Inquiry Monographs 16. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar: Handbook in Generative Syntax*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.

Rizzi, Luigi. 2006. On the Form of Chains: Criterial Positions and ECP effects. In *Wh-Movement: Moving On*, ed. Lisa Lai-Shen Cheng and Norbert Corver, 97–133. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2010. On Some Properties of Criterial Freezing. In *The Complementizer Phase: Subjects and Operators*, ed. Phoevos Panagiotidis, 17–32. Oxford: Oxford University Press.

Rizzi, Luigi, and Ur Shlonsky. 2007. Strategies of Subject Extraction. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, ed. Uli Sauerland and Hans Martin Gärtner, 115–160. Berlin: Mouton de Gruyter.

Roberts, Ian, and Anna Roussou. 2003. *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Romero, Maribel. 1997. Problems for a Semantic Account of Scope Reconstruction. In *Proceedings of the 1997 Tübingen Workshop on Reconstruction*, ed. Graham Katz, Shin-Sook Kim, and Haike Winhart, 127–153.

Rooryck, Johan. 1994. On two types of underspecification: Towards a feature theory shared by syntax and phonology. *Probus* 6:207–233.

Rooryck, Johan. 2000. *Configurations of Sentential Complementation*. London: Routledge.

Rooryck, Johan. 2003. The morphosyntactic structure of articles and pronouns in Dutch. In *Germania et alia. A linguistic webschrift for Hans den Besten*, ed. Jan Koster and Henk van Riemsdijk, 1–13. Available at `http://odur.let.rug.nl/koster/DenBesten/contents.htm`.

Rosenbaum, Peter S. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: MIT Press.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Ross, John Robert. 1969. Guess who? In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, ed. Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan, 252–286. Chicago Linguistic Society, University of Chicago, Chicago, Illinois.

Ross, John Robert, and David M. Perlmutter. 1970. Relative Clauses with Split Antecedents. *Linguistic Inquiry* 1:350.

Roussou, Anna. 2010. Selecting complementizers. *Lingua* 120:582–603.

Rullmann, Hotze. 1995. Maximality in the Semantics of Wh-Constructions. Doctoral Dissertation, University of Massachusetts Amherst.

Ruys, Eddy. 2011. Semantic Reconstruction and the Interpretation of Chains. In *Proceedings of Sinn und Bedeutung 15*, ed. Ingo Reich, Eva Horch, and Dennis Pauly, 515–529. Saarbrücken, Germany: Universaar – Saarland University Press.

Safir, Kenneth. 1999. Vehicle Change and Reconstruction in A-bar Chains. *Linguistic Inquiry* 30:587–620.

Salzmann, Martin. 2006. Resumptive Prolepsis. A study in indirect A'-dependencies. Doctoral Dissertation, Universiteit Leiden. LOT Dissertation Series 136.

Sasse, Hans-Jürgen. 1993. Syntactic Categories and Subcategories. In *Syntax. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 646–686. Berlin: De Gruyter.

Sauerland, Uli. 1998. The Meaning of Chains. Doctoral Dissertation, Massachusetts Institute of Technology.

Sauerland, Uli. 2003. Unpronounced heads in relative clauses. In *The Interfaces: Deriving and interpreting omitted structures*, ed. Kerstin Schwabe and Susanne Winkler, Linguistik Aktuell/Linguistics Today 61, 205–226. Amsterdam/Philadelphia: John Benjamins.

Schachter, Paul. 1973. Focus and relativization. *Language* 49:19–46.

Schippers, Ankelien. 2006. Variation in the CP-domain: Long-distance dependencies in the Dutch dialects. Ms., Meertens Instituut and Rijksuniversiteit Groningen.

Schmitt, Christina. 2000. Some Consequences of the Complement Analysis for Relative Clauses, Demonstratives and the Wrong Adjectives. In *The Syntax of Relative Clauses*, ed. Artemis Alexiadou, Paul Law, André Meinunger, and Chris Wilder, Linguistik Aktuell/Linguistics Today 32, 309–348. Amsterdam: John Benjamins.

Schoorlemmer, Erik. 2009. Agreement, Dominance and Doubling. Doctoral Dissertation, Universiteit Leiden. LOT Dissertation Series 225.

Sells, Peter. 1984. Syntax and Semantics of Resumptive Pronouns. Doctoral Dissertation, University of Massachusetts Amherst.

Sharvit, Yael. 1999a. Connectivity in specificational sentences. *Natural Language Semantics* 7:299–339.

Sharvit, Yael. 1999b. Functional Relative Clauses. *Linguistics and Philosophy* 22:447–478.

Sharvit, Yael, and Elena Guerzoni. 2003. Reconstruction and its problems. In *Proceedings of the 14th Amsterdam Colloquium*, ed. Paul Dekker and Robert van Rooy, 205–210. Universiteit van Amsterdam.

Siemund, Peter. 2008. *Pronominal Gender in English. A Study of English Varieties from a Cross-Linguistic Perspective*. London: Routledge.

Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. In *Grammatical Categories in Australian Languages*, ed. Robert M.W. Dixon, 112–171. Canberra: Australian Institute of Aboriginal Studies & New Jersey: Humanities Press.

Smith, Carlota S. 1964. Determiners and Relative Clauses in a Generative Grammar of English. *Language* 40:37–52.

Smits, Rik. 1988. The Relative and Cleft Constructions of the Germanic and Romance Langauges. Doctoral Dissertation, Katholieke Universiteit Brabant.

Sportiche, Dominique. 2006. Reconstruction, Binding and Scope. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume 4, chapter 54, 35–93. Oxford: Blackwell.

Sportiche, Dominique. 2008. Inward Bound: Splitting the wh-paradigm and French relative *qui*. Ms., UCLA.

Sportiche, Dominique. 2011. French Relative *Qui*. *Linguistic Inquiry* 42:83–124.

Stepanov, Arthur, and Penka Stateva. 2006. Successive cyclicity as residual wh-scope marking. *Lingua* 116:2107–2153.

Sternefeld, Wolfgang. 2001. Semantic vs. Syntactic Reconstruction. In *Linguistic Form and its Computation*, ed. Christian Rhorer, Antje Rossdeutscher, and Hans Kamp, 145–182. Stanford: CSLI Publications.

Sturm, Arie. 1996. Over functionele projecties. *Nederlandse Taalkunde* 1:191–206.

Suñer, Margarita. 2001. The Puzzle of Restrictive Relative Clauses with Conjoined DP Antecedents. In *Features and Interfaces in Spanish and French: Essays in Honor of Heles Contreras*, ed. Julia Herschensohn, Enrique Mallén, and Karen Zagona, 267–278. John Benjamins.

Svenonius, Peter, Gillian Catriona Ramchand, Michal Starke, and Knut

Tarald Taraldsen (eds.). 2009. Nordlyd (Tromsø University Working Papers on Language and Linguistics) 36.1: Special issue on Nanosyntax. University of Tromsø. Available at `http://septentrio.uit.no/index.php/nordlyd`.

Szabolcsi, Anna. 1987. Functional Categories in the Noun Phrase. In *Approaches to Hungarian 2*, ed. István Kenesei, 167–190. Szeged: JATE.

Szabolcsi, Anna. 1994. The noun phrase. In *The Syntactic Structure of Hungarian*, ed. Ferenc Kiefer and Katalin É. Kiss, Syntax and Semantics 27, 179–274. San Diego: Academic Press.

Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics of scope taking. *Natural Language Semantics* 1:235–284.

Taraldsen, Knut Tarald. 2001. Subject Extraction, the Distribution of Expletives, and Stylistic Inversion. In *Subject Inversion in Romance and the Theory of Universal Grammar*, ed. Aafke Hulk and Jean-Yves Pollock, Oxford Studies in Comparative Syntax, 163–181. Oxford: Oxford University Press.

Thoms, Gary. 2010. Syntactic reconstruction and Scope Economy. Handout of a talk presented at GLOW 33, Wroclaw, Poland.

Toebosch, Annemarie. 2007. A feature-geometry of gender and animacy: The case of deficient pronouns in Dutch. *MorphOn* 1–36.

Uriagereka, Juan. 1995. Aspects of the Syntax of Clitic Placement in Western Romance. *Linguistic Inquiry* 26:79–123.

Vangsnes, Øystein Alexander. 2005. Microparameters for Norwegian *wh*-grammars. *Linguistic Variation Yearbook* 5:187–226.

Vergnaud, Jean-Roger. 1974. French Relative Clauses. Doctoral Dissertation, Massachusetts Institute of Technology.

Vikner, Sten. 1991. Relative *der* and other C° elements in Danish. *Lingua* 84:109–136.

Vikner, Sten. 1994. Finite Verb Movement in Scandinavian Embedded Clauses. In *Verb Movement*, ed. Norbert Hornstein and David Lightfoot, 117–147. Cambridge: Cambridge University Press.

Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.

de Vries, Mark. 2002. The Syntax of Relativization. Doctoral Dissertation, Universiteit van Amsterdam. LOT Dissertation Series 53.

de Vries, Mark. 2004. Head-Internal Relative Clauses in Dutch? In *Linguistics in the Netherlands 2004*, ed. Leonie Cornips and Jenny Doetjes, 193–204. Amsterdam/Philadelphia: John Benjamins.

de Vries, Mark. 2006. Possessive Relatives and (Heavy) Pied Piping. *Journal of Comparative Germanic Linguistics* 9:1–52.

Wasow, Thomas, Geoffrey Nunberg, and Ivan A. Sag. 1984. Idioms: An Interim

Report. In *Proceedings of the XIIIth International Congress of Linguists*, ed. Shiro Hattori and Kazuko Inoue, 102–115. The Hague: CIPL.

Webelhuth, Gert. 2011. Capturing Collocations and Idioms in Relative Clauses without Literal Reconstruction. Handout of a talk presented at the workshop 'Reconstruction Effects in Relative Clauses', Zentrum für Allgemeine Sprachwissenschaft (ZAS), Berlin, July 2011.

Weerman, Fred, and Jacqueline Evers-Vermeul. 2002. Pronouns and Case. *Lingua* 112:301–338.

Weiß, Helmut. 2002. Three Types of Negation. A Case Study in Bavarian. In *Syntactic Microvariation*, ed. Sjef Barbiers, Leonie Cornips, and Suzanne van der Kleij, 305–332. Meertens Institute Electronic Publications in Linguistics (MIEPiL). Available at `http://www.meertens.knaw.nl/books/synmic/`.

Westergaard, Marit, and Øystein Alexander Vangsnes. 2005. *Wh*-questions, V2 and the Left Periphery of Three Norwegian Dialect types. *Journal of Comparative Germanic Linguistics* 8:117–158.

Wexler, Kenneth, and Peter W. Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Wilder, Christopher. 1991. Tough movement constructions. *Linguistische Berichte* 132:115–132.

Wiltschko, Martina. 1998. On the Syntax and Semantics of (Relative) Pronouns and Determiners. *Journal of Comparative Germanic Linguistics* 2:143–181.

Wiltschko, Martina. 2002. The Syntax of Pronouns: Evidence from Halkomelem Salish. *Natural Language & Linguistic Theory* 20:157–195.

Zanuttini, Raffaella, and Paul Portner. 2003. Exclamative Clauses: At the Syntax-Semantics Interface. *Language* 79:39–81.

Zeijlstra, Hedde. 2010. There is only one way to agree. Handout of a talk presented at GLOW 33, Wroclaw, Poland.

Zwart, Jan-Wouter. 1993. Dutch Syntax: A Minimalist Approach. Doctoral Dissertation, Rijksuniversiteit Groningen.

Zwart, Jan-Wouter. 2000. A Head Raising Analysis of Relative Clauses in Dutch. In *The Syntax of Relative Clauses*, ed. Artemis Alexiadou, Paul Law, André Meinunger, and Chris Wilder, Linguistik Aktuell/Linguistics Today 32, 349–386. Amsterdam: John Benjamins.

Zwart, Jan-Wouter, and Eric Hoekstra. 1997. Weer functionele projecties. *Nederlandse Taalkunde* 2:121–132.

Samenvatting in het Nederlands (Summary in Dutch)

Dit proefschrift geeft een gedetailleerde beschrijving en analyse van de morfosyntactische microvariatie in restrictieve relatiefzinnen en gerelateerde constructies (zoals vraagwoordvragen) in het Nederlands. Een *relatiefzin*, ofwel een *betrekkelijke bijzin*, is een ondergeschikte zin die een nomen modificeert – het *hoofd* of *antecedent* van de relatiefzin. Een relatiefzin wordt (vaak) ingeleid door een betrekkelijk voornaamwoord – een *relatiefpronomen* – dat verwijst naar het antecedent, en qua vorm ook afhankelijk is van het antecedent. Een Standaardnederlandse relatiefzin verschijnt altijd rechts van het antecedent dat het modificeert, en moet ingeleid worden door een relatiefpronomen. Het gerelativiseerde element – ik laat voor nu in het midden of dat alleen het relatiefpronomen of het relatiefpronomen *plus* het antecedent van de relatiefzin is – ondergaat verplaatsing vanuit een argumentspositie naar een positie in de linkerperiferie van de zin – het *CP domein* (namelijk SpecCP, een *A-bar* positie) – en laat daarbij een kopie achter in zijn oorspronkelijke positie – de *extractieplaats*. In het Standaardnederlands wordt deze kopie niet fonologisch gerealiseerd. Dit is geïllustreerd in (1), waarbij de extractieplaats is weergegeven als _.

(1) de man [$_{RC}$ die ze _ geroepen hebben] *Standaardnederlands*

Relativisatie kan ook plaatsvinden vanuit een ingebedde zin, wat resulteert in *lange-afstandsrelativisatie* ofwel *lange relatiefzinnen*. Dit is geïllustreerd in (2).

(2) de man [$_{RC}$ die ik denk [dat ze _ geroepen hebben]] *Standaardnederlands*

Volgens de traditionele aanname maakt verplaatsing van het gerelativiseerde element een tussenlanding in de linkerperiferie van de ingebeddde zin – *successief-cyclische verplaatsing* via SpecCP – en laat daarbij een kopie achter. Een dergelijke syntactische verplaatsingsketen wordt ook wel *A-bar verplaatsingsketen* genoemd.

Zowel bestaande data – de SAND data: Barbiers et al. (2005, 2008a) – als nieuwe data – verworven uit grootschalige online questionnaires (de Meertens Panel Questionnaires) – laten een grote mate van variatie zien in lange relatiefzinnen (en lange vraagwoordvragen) in het Nederlandssprekende taalgebied. Die variatie betreft vooral de vorm van de elementen die de hogere en de lagere zin introduceren. De constructie die centraal staat in dit proefschrift is *verdubbeling* van een (relatief)pronomen in lange relatiefzinnen (en verdubbeling van een pronomen in vraagwoordvragen), zoals geïllustreerd in (3) voor identieke verdubbeling van het pronomen *die*. Verdubbeling van pronomina in lange A-bar ketens heeft geen duidelijke geografische distributie en komt voor in informeel gesproken Nederlands.

(3) de man [$_{RC}$ **die** ik denk [**die** ze _ geroepen hebben]]
informeel gesproken Nederlands

Dit proefschrift laat zien dat een goed begrip van microvariatie in relatiefzinnen en gerelateerde constructies, en van *verdubbeling* in het bijzonder, nieuwe inzichten verschaft in de structuur van (lange) A-bar ketens (hoofdstuk 2), de syntaxis van relatiefzinnen, de structuur van de linkerperiferie (hoofdstuk 3), en de aard van (relatief)pronomina en hun relatie met voegwoorden (hoofdstuk 4). Meer in het bijzonder draagt deze studie bij aan een nieuwe, overkoepelende analyse van verdubbeling in lange relatiefzinnen en verdubbeling in lange vraagwoordvragen, aan een herinvoering van (een specifieke variant van) de traditionele hoofdexterne analyse van relatiefzinnen, en aan een nieuwe implementatie van een onderspecificatie analyse van zogenaamde *multifunctionele pronomina*: pronomina die meer dan één functie kunnen hebben en in meer dan één syntactische configuratie kunnen voorkomen. Hieronder bespreek ik kort de inhoud van elk hoofdstuk in dit proefschrift.

Hoofdstuk 2 geeft een analyse van verdubbeling in lange A-bar ketens op basis van nieuw verkregen empirische data over verdubbeling van pronomina in lange relatiefzinnen en verdubbeling van pronomina in lange (ingebedde) vraagwoordzinnen in het Nederlands. De belangrijkste verdubbelingspatronen die in dit hoofdstuk centraal staan, zijn gegeven in (4) voor een lange relatiefzin met een non-neutrum menselijk antecedent, en in (5) voor een lange vraagwoordvraag die naar een persoon vraagt.

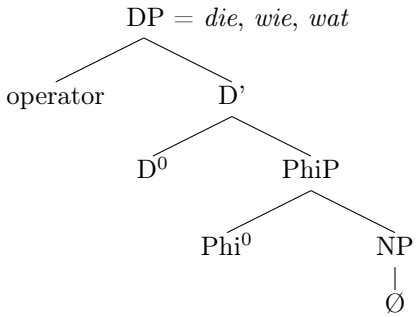
(4) de man <**die**/**wie**> ik denk <**die**/**wie**> het gedaan heeft

(5) <**wie**/**wat**> denk je <**die**/**wie**> het gedaan heeft?

Ik neem aan dat alle lange A-bar ketens afgeleid zijn door successief-cyclische verplaatsing via SpecCP van (een deel van) het *A-bar pronomen* (interrogatief- of relatiefpronomen), en dat verdubbeling het resultaat is van de uitspelling van meer dan één (deel van een) kopie van het A-bar pronomen in de fonologische component (PF). Meer in het bijzonder laat ik zien dat de interne structuur

van A-bar pronomina een *operator* bevat die zich in de specificeerderpositie van het pronomen bevindt (zie o.a. Szabolcsi 1994), zoals geïllustreerd in (6). Het is de operator in A-bar pronomina die verplaatsing naar de linkerperiferie afdwingt.

(6) De structuur van A-bar pronomina



Op het moment in de derivatie dat het A-bar pronomen het ingebedde CP domein heeft bereikt, ontstaan er twee mogelijkheden: ofwel het gehele pronomen verplaatst naar het hogere CP domein, ofwel alleen de operator zelf verplaatst naar het hogere CP domein. Als het pronomen verplaatst, is het resultaat de verplaatsingsketen in (7).

(7) $[_{CP}$ pronomen$_1$ ... $[_{CP}$ pronomen$_1$ ... pronomen$_1$ ...]]
verplaatsing van het pronomen

Onder normale omstandigheden mag alleen de hoogste kopie in een verplaatsingsketen uitgespeld (fonologisch gerealiseerd) worden. Alle andere kopieën moeten uitgewist of gedeleerd (fonologisch niet gerealiseerd) worden op PF (PF-deletie). Dit is geïllustreerd in (8a) voor de verplaatsingsketen van het A-bar pronomen in (7); doorstreping geeft PF-deletie aan. Onder bepaalde omstandigheden kan de kopie van het pronomen in de lagere SpecCP ook uitgespeld worden, naast de uitspelling van het hoofd van de verplaatsingsketen (zoals voorgesteld door Nunes 2004). Dit resulteert in *identieke verdubbeling*, zoals is te zien in (8b).

(8) a. $[_{CP}$ **pronomen**$_1$ ... $[_{CP}$ ~~pronomen$_1$~~ ... ~~pronomen$_1$~~ ...]]
geen verdubbeling

b. $[_{CP}$ **pronomen**$_1$ ... $[_{CP}$ **pronomen**$_1$ ... ~~pronomen$_1$~~ ...]]
identieke verdubbeling

Als in plaats van het pronomen alleen de operator verplaatst – dat wil zeggen, *subextractie* van de operator, ofwel de afwezigheid van *pied piping* van het gehele pronomen (subextractie en *pied piping* zijn twee zijden van dezelfde medaille) – ziet de syntactische verplaatsingsketen eruit als in (9).

(9) $[_{CP}$ operator$_1$ ... $[_{CP}$ pronomen$_1$ ... pronomen$_1$ ...]]
subextractie van de operator

De geëxtraheerde operator wordt uitgespeld in de hogere SpecCP. Het pronomen in de lagere SpecCP, waaruit de operator geëxtraheerd is, moet ook uitgespeld worden, zodat de informatie die besloten ligt in het pronomen niet verloren gaat. Dat wil zeggen, de informatie in het pronomen is niet zichtbaar als het pronomen geledeerd wordt op PF.

Subextractie van de operator is feitelijk een schending van localiteitsprincipes zoals de *Condition on Extraction Domain* (CED, Huang 1982) en het *Freezing Principle* (Wexler en Culicover 1980), volgens welke een constituent die al verplaatsing heeft ondergaan geen verdere verplaatsing vanuit die constituent toestaat. Ik stel voor dat door uitspelling van het gehele pronomen waaruit de operator geëxtraheerd is, een schending van dergelijke localiteitsprincipes teniet wordt gedaan (zie ook Van Craenenbroeck en Van Koppen 2008). Deze manier om een derivatie te 'redden' noem ik *rescue by PF spell out* – de logische tegenhanger van *rescue by PF deletion* (zie o.a. Bošković 2011). Onder de aanname dat A-bar pronomina de uitspelling van frases zijn (zie o.a. Weerman en Evers-Vermeul 2002, Barbiers, Koeneman en Lekakou 2009, en zie (6)), zal de uitspelling van een pronomen altijd de uitspelling van een operator *bevatten*. Het resultaat hiervan is dat de kopie van het pronomen in het ingebedde CP domein (zoals in (9)) altijd uitgespeld wordt als een pronomen. Subextractie en dubbele uitspelling is abstract geïllustreerd in (10).

(10) $[_{CP}$ **operator**$_1$ ... $[_{CP}$ **pronomen**$_1$ ... ~~pronomen$_1$~~ ...]]
niet-identieke verdubbeling

Een voorbeeld van deze constructie is het verdubbelingspatroon *wat-die/wie* in (5); de geëxtraheerde operator wordt uitgespeld als *wat*, omdat *wat* het meest ondergespecificeerde A-bar pronomen in het Nederlands is (zie o.a. Postma 1994, Bennis 1995). Merk op dat het verdubbelingspatroon *wat-die/wie* niet voorkomt in relatiefzinnen met een non-neutrum menselijk antecedent (zie (4)), omdat zulke relatiefzinnen niet ingeleid kunnen worden door het pronomen *wat*. Dit patroon lijkt echter wel voor te komen in relatiefzinnen met een neutrum menselijk antecedent, zoals *meisje* (bijvoorbeeld *het meisje wat ik denk die het gedaan heeft*). Operator verplaatsing en dubbele uitspelling lijkt dus niet voorbehouden te zijn aan lange vraagwoordvragen, maar lijkt een meer algemene strategie te zijn voor de vorming van lange A-bar verplaatsingsketens.

De specifieke verdubbelingspatronen in relatiefzinnen en vraagwoordvragen, zoals gegeven in (4) en (5), kunnen het beste verklaard worden door de kenmerkspecificaties van de relevante A-bar pronomina en de verschillende lexicalisatie (uitspelling) opties. Dat wil zeggen, de specifieke distributie van de relevante A-bar pronomina is afhankelijk van de kenmerken die ze kunnen lexicaliseren: syntactisch geslacht (neutrum/non-neutrum) en/of semantische 'animacy' (ruwweg menselijk/niet-menselijk). Een grammaticamodel waarin morfologie

pas plaatsvindt *na* de syntaxis – in een dergelijk model wordt in de morfologische component bepaald hoe de bundels abstracte kenmerken (waarop de syntaxis opereert) het best gelexicaliseerd kunnen worden – kan een verklaring geven voor de observatie dat bepaalde pronomina inwisselbaar zijn in sommige contexten (voor sommige sprekers). Zulke pronomina kunnen even goed een bepaalde syntactische structuur of bundel kenmerken lexicaliseren. Zo kunnen de A-bar pronomina *die* en *wie* even goed het kenmerk [menselijk] lexicaliseren, met als resultaat dat ze inwisselbaar zijn in de hogere en lagere zin van een lange relatiefzin met een menselijk antecedent, en in de lagere zin van een lange vraagwoordvraag die naar een persoon vraagt; zie de verdubbelingspatronen in (4) en (5). De keuze voor welk pronomen een bepaalde structuur lexicaliseert, is bovendien afhankelijk van de aard van de zin: relatiefzin of vraagwoordvraag. Dat wil zeggen, een *d*-pronomen kan niet een vraagwoordvraag introduceren vanwege de eis dat vraagwoordzinnen geïntroduceerd moeten worden door *w*-woorden.

De meest prominente analyses van relatiefzinnen in de huidige literatuur – zogenaamde *Hoofd Interne Analyses* (HIA's) van relatiefzinnen – nemen aan dat het antecedent van de relatiefzin (samen met het relatiefpronomen) zijn oorsprong vindt *in* de relatiefzin en vervolgens verplaatst naar de linkerperiferie van de zin. Daar wordt het ofwel zelf het hoofd van de relatiefzin (*raising*), ofwel het wordt gedeleerd onder identiteit met een hoofd dat buiten de relatiefzin staat (*matching*). **Hoofdstuk 3** laat zien dat (de in hoofdstuk 2 voorgestelde analyse van) verdubbeling in lange relatiefzinnen het makkelijkst te combineren is met een analyse van relatiefzinnen waarin het hoofd van de relatiefzin niet in de argumentspositie in de relatiefzin zelf basis-gegenereerd wordt: een *Hoofd Externe Analyse* (HEA) van relatiefzinnen. Dat wil zeggen, onder een analyse van verdubbeling in termen van de PF-uitspelling van meerdere kopieën, voorspelt een HIA van relatiefzinnen dat het hoofd van de relatiefzin ook meerdere keren uitgespeld wordt in verdubbelingsconstructies, *quon non*. Dat deze voorspelling niet uitkomt, is geïllustreerd in (11) voor een *raising* analyse van relatiefzinnen à la Kayne (1994).[1]

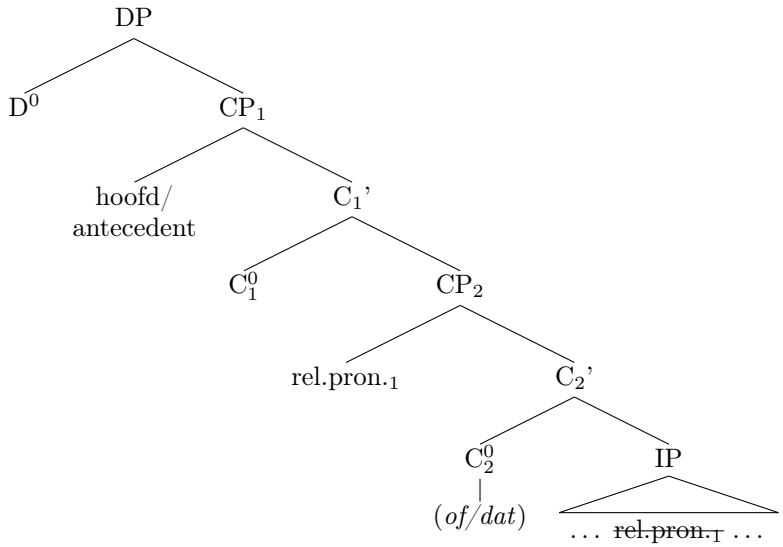
(11) * de [[**man**$_2$ **die** ~~man$_2$~~]$_1$ ik denk [[**die man**]$_1$ ze ~~[die man]$_1$~~ geroepen hebben]]

Naar aanleiding van de vaststelling dat een HIA van relatiefzinnen niet eenvoudig een verklaring voor de verdubbelingsfeiten kan geven, stelt hoofdstuk 3 een specifieke implementatie van een HEA van relatiefzinnen voor. Deze analyse gaat uit van een D-complement hypothese (zie o.a. Smith 1964, Kayne 1994) – die stelt dat de externe determineerder de relatiefzin (CP) als zijn complement neemt – en van de hypothese dat het CP domein is opgesplitst in verschillende projecties (zie o.a. Van Craenenbroeck 2004, 2010 voor het Nederlands). Het hoofd van de relatiefzin wordt basis-gegenereerd in de hoogste SpecCP positie

[1] Een * voor een zin geeft aan dat die zin ongrammaticaal is.

van de relatiefzin zelf, terwijl het relatiefpronomen van zijn basispositie verplaatst naar de lagere SpecCP positie (zie ook Schmitt 2000, Aoun en Li 2003). Dit is geïllustreerd in (12). Strikt genomen is de analyse in (12) een HIA omdat het hoofd van de relatiefzin basis-gegenereerd wordt *in* de relatiefzin (CP) zelf. Toch beschouw ik de structuur in (12) als een HEA, aangezien het hoofd van de relatiefzin niet in de argumentspositie in de relatiefzin basis-gegenereerd wordt en niet verplaatst naar de linkerperiferie.

(12) De structuur van restrictieve relatiefzinnen in het Nederlands



Deze analyse kan niet alleen een verklaring geven voor de verdubbelingsdata, maar ook voor alle variatie in de linkerperiferie van Nederlandse relatiefzinnen, in het bijzonder dubbel gevulde Comp data. Dubbel gevulde Comp is het fenomeen waarbij de linkerperiferie van een ingebedde zin geïntroduceerd wordt door een pronomen *en* een (of meer) voegwoord(en).

Het resterende deel van hoofdstuk 3 evalueert de HEA van relatiefzinnen, en laat zien dat de keuze voor een HEA of een HIA van relatiefzinnen zeker niet triviaal is, ook al zijn HIA's (*raising* of *matching*) de standaard analyses van relatiefzinnen in de huidige literatuur. De HEA blijkt het in veel opzichten beter te doen dan HIA's. Zo kan een HEA bijvoorbeeld beter een verklaring geven voor verschillen in naamval tussen het hoofd van de relatiefzin en het relatiefpronomen. Alleen de aanwezigheid van *connectiviteitseffecten* tussen (materiaal in) het hoofd van de relatiefzin en de extractieplaats in de relatiefzin lijkt een sterk argument tegen de HEA van relatiefzinnen te zijn. Connectiviteitseffecten zijn effecten waarbij een constituent geïnterpreteerd wordt in een positie die verschilt van zijn oppervlaktepositie. In het geval van relatiefzinnen betekent dit dat het hoofd van de relatiefzin geïnterpreteerd wordt in de extractieplaats

in de relatiefzin. Als connectiviteitseffecten het resultaat zijn van *syntactische reconstructie* – dat wil zeggen, de activatie van een lagere kopie in de semantische component (LF) (zie o.a. Chomsky 1993, Fox 1999b) – moet het hoofd van de relatiefzin dus in argumentspositie in de relatiefzin zelf basis-gegenereerd zijn. Ik laat echter zien – in overeenstemming met huidige literatuur (zie bijvoorbeeld Sharvit 1999a, Sharvit en Guerzoni 2003, Cecchetto 2005, en Heycock 2005) – dat connectiviteitseffecten niet altijd verklaard kunnen worden door syntactische reconstructie. Anders gezegd, reconstructie zonder verplaatsing of kopieën lijkt sowieso nodig te zijn, zowel in relatiefzinnen als in andere configuraties, zoals linksdislocatie constructies of specificationele 'pseudocleft' constructies. Ik toon dus aan dat connectiviteitseffecten niet altijd een betrouwbare diagnostiek zijn voor verplaatsing, en dat dientengevolge de aanwezigheid van connectiviteitseffecten in relatiefzinnen onvoldoende evidentie geeft voor de ondersteuning van een HIA van relatiefzinnen.

Het is welbekend dat veel talen gebruik maken van relatiefpronomina die ook in andere pronomenparadigma's gebruikt worden, zoals het demonstratieve of het interrogatieve pronomenparadigma. Relatiefpronomina hebben dus vaak meer dan één functie en kunnen voorkomen in meer dan één configuratie (*multifunctionele pronomina*). Zo kan *wat* in het Nederlands naast relatiefpronomen (*alles wat hij gedaan heeft*) onder andere ook gebruikt worden als vraagwoord (*wat heeft hij gedaan?*) en als indefiniet of onbepaald pronomen (*hij heeft wat gedaan*). Het eerste deel van **hoofdstuk 4** behandelt de aard van zulke multifunctionele pronomina – en de aard van relatiefpronomina in het bijzonder – en vertrekt vanuit de nulhypothese dat een multifunctioneel pronomen morfosyntactisch en semantisch *ondergespecificeerd* is; we hebben dus te maken met één ondergespecificeerd lexicaal item, in plaats van met meerdere lexicale items die toevallig homofoon zijn. Als gevolg van de onderspecificatie van een multifunctioneel pronomen kan het in verschillende syntactische omgevingen voorkomen en wordt een deel van de betekenis van het pronomen contextueel of configurationeel bepaald. Zo laat ik voor demonstratieve pronomina bijvoorbeeld in detail zien dat distale demonstratieven (*die* en *dat*) ondergespecificeerd zijn voor *locatie*, terwijl proximale demonstratieven (*deze* en *dit*) gespecificeerd zijn voor locatie (als proximaal), zoals eerder ook voorgesteld door Rooryck (2003). Een consequentie van deze kenmerkspecificatie is dat alleen distale demonstratieven kunnen fungeren als relatiefpronomina.

Net als relatiefpronomina corresponderen finiete declaratieve voegwoorden vaak met pronomina uit het demonstratieve of interrogatieve pronomenparadigma. Het Nederlandse voegwoord *dat* correspondeert met het distale demonstratieve neutrum pronomen. Hoewel ik niet zo ver ga en aanneem dat voegwoord *dat* en pronomen *dat* hetzelfde lexicale item zijn in het Nederlands, neem ik wel aan dat ze diachroon gerelateerd zijn. Het is precies het meest ondergespecificeerde *d-pronomen* dat als bron dient voor grammaticalisatie in een voegwoord (in overeenstemming met de nulhypothese); *neutrum* wordt gezien

als de onderspecificatie voor geslacht en *distaal* is de onderspecificatie voor locatie.

Een recentelijk populaire visie onder taalkundigen is het idee dat complementatie eigenlijk relativisatie is (o.a. Arsenijević 2009) en dat voegwoorden eigenlijk relatiefpronomina zijn (Kayne 2008, 2010). In het laatste deel van hoofdstuk 4 pleit ik sterk tegen dit idee door te laten zien dat het tegen allerlei theoretische en empirische problemen aanloopt, waaronder het onvermogen om een adequate verklaring te geven voor de (variatie in) dubbel gevulde COMP patronen in het Nederlands. Bovendien bespreek ik een casusstudie die laat zien dat het noodzakelijk is om relatiefpronomina te onderscheiden van voegwoorden. Deze casusstudie richt zich op Zuidnederlandse dialecten die gebruik maken van het element *die* in de meest ingebedde zin van (lange) *subject* relatiefzinnen: 'speciaal *die*'. Met name op basis van de observatie dat er een positieve correlatie is tussen de distributie van speciaal *die* en de distributie van voegwoordvervoeging – het fenomeen waarbij een finiet declaratief voegwoord congruentie vertoont met het subject van de zin die het introduceert – laat ik zien dat speciaal *die* het best geanalyseerd kan worden als een congruerende variant van het voegwoord, in plaats van als een relatiefpronomen.

Hoofdstuk 5 geeft een samenvatting en presenteert een overzicht van de meest belangrijke empirische en theoretische bijdragen van dit proefschrift. Deze zijn hieronder nog een keer op een rij gezet.

- een gedetailleerd overzicht van de (grenzen aan) variatie in verdubbeling in lange A-bar verplaatsingsketens in het Nederlands
- een nieuwe, overkoepelende analyse van verdubbeling in lange relatiefzinnen en verdubbeling in lange vraagwoordzinnen
- uitspelling van een constituent op PF kan een ongrammaticale stap in een derivatie redden
- de syntaxis kan de locus zijn van microvariatie
- de herinvoering van (een specifieke variant van) de traditionele Hoofd Externe Analyse van relatiefzinnen
- connectiviteitseffecten zijn niet altijd een betrouwbare diagnostiek voor verplaatsing
- een multifunctioneel pronomen is een morfosyntactisch en semantisch ondergespecificeerd lexicaal item
- niet alle ingebedde zinnen zijn relatiefzinnen
- speciaal *die* is een voegwoord, niet een relatiefpronomen

Curriculum Vitae

Eefje Boef werd op 21 december 1983 geboren in Utrecht. Zij behaalde haar gymnasiumdiploma aan het Stedelijk Gymnasium te Arnhem. Na haar middelbare school begon zij in 2003 met de studie *Nederlandse taal en cultuur* aan de Universiteit Utrecht. Na het cum laude behalen van haar bachelordiploma (met specialisatierichting taalkunde) in 2006, startte zij met de tweejarige onderzoeksmaster *Linguistics: the Study of the Language Faculty* (Prestige Master) aan het Utrecht Institute of Linguistics OTS (UiL OTS) van de Universiteit Utrecht. Tijdens deze masteropleiding was zij gedurende een aantal maanden werkzaam als assistent-onderzoeker aan het Meertens Instituut in Amsterdam (KNAW), waar zij meewerkte aan de afronding van het tweede deel van de SAND (*Syntactische Atlas van de Nederlandse Dialecten*). Na het cum laude afronden van de onderzoeksmaster in de zomer van 2008, begon zij in september van datzelfde jaar als Onderzoeker in Opleiding (OiO) in het Edisyn (*European Dialect Syntax*) project aan het Meertens Instituut. Tijdens haar promotietraject verbleef zij twee keer voor een langere periode in het buitenland. Van januari tot en met juni 2010 was zij met een (Yggdrasil) beurs van de Noorse Research Council als *visiting researcher* verbonden aan CASTL (*Center for Advanced Study in Theoretical Linguistics*) aan de Universiteit van Tromsø, Noorwegen. Vervolgens ging zij in januari 2011 voor drie maanden naar New York, waar zij haar onderzoek voortzette bij de taalkunde departementen van *New York University* (NYU) en *The Graduate Center* van *The City University of New York* (CUNY). Dit proefschrift is het resultaat van haar werkzaamheden in binnen- en buitenland gedurende haar aanstelling als OiO in het Edisyn project. Sinds september 2012 is zij werkzaam als onderzoeker aan het *Zentrum für Allgemeine Sprachwissenschaft* (ZAS) in Berlijn.