

Dependencies across Phases

*From Sequence of Tense to Restrictions on
Movement*

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

Phone: +31 30 253 6006
Fax: +31 30 253 6406
e-mail: lot@let.uu.nl
<http://www.lotschool.nl>

Cover illustration by Felix Dankevitch

ISBN 978-90-78328-43-8
NUR 616

Copyright © 2007: Olga Khomitsevich. All rights reserved.

This dissertation is typeset using \LaTeX .

Dependencies across Phases
*From Sequence of Tense to Restrictions on
Movement*

Afhankelijkheden over Fasen heen
*Van opeenvolging van tijden tot beperkingen op
verplaatsing*

(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. J. C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op vrijdag 11 januari 2008
des middags te 12.45 uur

door

Olga Khomitsevich

geboren op 24 mei 1981 te Sint-Petersburg, Russische Federatie

Promotoren: Prof. dr. E.J. Reuland
Prof. dr. T.V. Chernigovskaya

Contents

Acknowledgements	vii
1 Introduction	1
1.1 What this dissertation is about	1
1.2 Outline of the data	1
1.3 The proposal	3
1.4 Grammatical properties of Russian	4
1.4.1 Temporal and aspectual system of Russian	5
1.4.2 Complement clauses in Russian	8
1.4.3 Wh-movement and wh-questions in Russian	11
1.5 Structure of the dissertation	12
2 General syntactic background	15
2.1 Clause structure	15
2.1.1 The C-T system	15
2.1.2 Phase theory	18
2.1.3 Pesetsky and Torrego’s theory of C-T dependency	21
2.1.4 Summary	35
2.2 Binding theory	36
2.2.1 Classical binding theory	36
2.2.2 Syntactic binding and logical syntax binding	38
2.2.3 A new version of BT	39
2.2.4 Binding and agreement	45
2.2.5 Summary	48
3 Sequence of Tense	49
3.1 An overview of the data	49
3.1.1 SOT: The problem	49
3.1.2 “Truly vacuous” past	52
3.1.3 Tenses in relative and adjunct clauses	52
3.1.4 Tenses embedded under future	54
3.1.5 SOT facts cross-linguistically	55
3.2 Theories of SOT	58
3.2.1 The nature of tenses	59

3.2.2	Accounts of SOT	63
3.3	Summary	80
4	The analysis of SOT in English and Russian	83
4.1	The nature of simultaneous past	83
4.1.1	Simultaneous past in SOT languages	84
4.1.2	Simultaneous past in Russian	90
4.1.3	Summary	104
4.2	SOT: The account	104
4.2.1	SOT in “SOT languages”	104
4.2.2	SOT in Russian	112
4.3	Remaining issues	114
4.3.1	Deriving past-shifted readings	114
4.3.2	Present under past in Russian and English	116
4.3.3	Complements of NPs	117
4.3.4	Embeddings under future	119
4.4	Summary	123
5	Long-distance wh-movement in Russian	125
5.1	Locality of movement	126
5.1.1	Island constraints	126
5.1.2	Theories of locality	127
5.1.3	Locality of movement and successive cyclicity in Phase Theory	130
5.2	Locality of movement in Russian	134
5.2.1	Russian data: the puzzle	134
5.2.2	Existing accounts	137
5.2.3	Criticism of the existing accounts	144
5.3	Long-distance movement in Russian: The account	149
5.3.1	Phases in Russian and successive-cyclic movement	149
5.3.2	Remaining issues	154
5.4	Summary	160
6	Conclusions	163
	References	167
	Samenvatting in het Nederlands	177
	Curriculum vitae	181

Acknowledgements

Many people helped me and supported me while I was writing this dissertation, and now it is time to say thank you. First of all, I thank my supervisors Eric Reuland (at Utrecht) and Tatiana V. Chernigovskaya (at St. Petersburg). Eric's ideas, criticism and comments were always extremely valuable, and he helped me to present my arguments in a thorough and logical way. I am deeply grateful, too, for his never-failing enthusiasm for and reassurance about my work. Tatiana Vladimirovna was my supervisor when I was an undergraduate student at St. Petersburg, and I must thank her for teaching me how to think like a linguist and introducing me to current problems in the field. And of course, I am grateful for her encouragement and her interest in my dissertation research and for her great help with the St. Petersburg part of my PhD duties.

I am indebted to many people who discussed my work with me, answered my questions and kindly sent me their work, including, but not limited to, Daniel Altshuler, Sergey Avrutin, Ora Matushansky, Philippe Schlenker, Natalia Slioussar, Arthur Stepanov, and Michael Yadroff (the names are listed alphabetically). I am particularly grateful to Øystein Nilsen and Olga Borik for reading an earlier version of this dissertation and providing many useful comments. (Needless to say, none of these people are in any way responsible for my mistakes.) For giving their judgments as native speakers, I thank Boban Arsenijevic, Anna Asbury, Jakub Dotlacil, Kevin Hawkins, Ania Kijak, Marjo van Koppen, Natasa Milesevic, Rick Nouwen, and Mirjam Rigterink. I am extremely grateful to all participants of my Russian questionnaires, as well as to all my Russian friends who I turned to with my Russian examples time and time again (Nadya Ditch deserves a special mention). A big thank you goes to Rick Nouwen for translating the Dutch summary of the dissertation on very short notice. Finally, I thank Felix Dankevitch for the beautiful cover illustration.

My research was carried out as part of the St. Petersburg-Utrecht PhD cooperation program, so it is appropriate here to thank the people who made the program possible: Sergey Avrutin, Tatiana V. Chernigovskaya, Martin Everaert, Eric Reuland, and Maaike Schoorlemmer. I also thank Rianne Giethoorn, Keetje van den Heuvel, Corrie van Os, and Maaike for

helping me to overcome the numerous bureaucratic difficulties connected with my trips to the Netherlands. It was not an easy task: every time I thought I now knew all there was to know about visas and residence permits, a new and unexpected problem arose and they helped me out.

My fellow PhD students and postdocs at UiL OTS made my stays in Utrecht a lot of fun. I thank them all: Anna Asbury, Fabian Battaglini, Natalie Boll-Avetisyan, Ivana Brasileiro, Bert Le Bruyn, Matteo Capelletti, Anna Chernilovskaya, Alexis Dimitriadis, Jakub Dotlacil, Gaetano Fiorin, Berit Gehrke, Gianluca Giorgolo, Nicole Grégoire, Nino Grillo, Sander van der Harst, Annemarie Kerkhoff, Cem Keskin, Ania Kijak, Marieke Kolkman, Arnout Koornneef, Marjo van Koppen, Huib Kranendonk, Joke de Lange, Tom Lentz, Shakuntala Mahanta, Marijana Marelj, Hannah de Mulder, Øystein Nilsen, Rick Nouwen, Min Que, Gyuri Rakosi, Mirjam Rigterink, Oren Sadeh-Leicht, Marieke Schouwstra, Anca Sevcenco, Natalia Slioussar, Giorgos Spathas, Kriszta Szendrői, Roberta Tedeschi, Christina Unger, Sharon Unsworth, Nada Vasic, Mario van de Visser, Evangelia Vlachou, Arjen Zondervan, Peter Zubkov and anyone else I might have forgotten. In particular, I thank Matteo, Marjo, Øystein and Annemarie for being great office-mates; another extra thank you goes to Matteo and Øystein for their help with \LaTeX . I also thank my housemates Cem, Natalia, Roberta, and Christina for making our shared house a comfortable home.

Going back in time (and space), I thank the Department of Theoretical Linguistics at St. Petersburg State University where I got my MA degree; even though I eventually chose to concentrate on syntax for my PhD, the choice was quite difficult, since I found so many courses on various branches of linguistics – from language acquisition to Native American languages – highly informative and inspiring.

Even further back, I feel like this is the right moment to thank everybody at my high school, the St. Petersburg Classical School 610. The influence of my school teachers and my fellow students has shaped my mind, and there I learnt to express my thoughts on paper clearly and coherently. I am especially aware of my debt to my language teachers, who helped me to discover linguistics and instilled a love for language in me: Natalia V. Chernukha, Elena N. Gracheva, Vanda P. Kazanskene, and Elena K. Yuzbashyan.

On a more informal note, I thank all my friends in Russia and around the world for keeping in touch via all possible electronic means, such as email, SMS, chat, and blogging. They never failed to provide the latest gossip, discussions of films and political debate that really kept me going. I am especially grateful to (alphabetically, as usual) Aleksey Andreev, Alena Buzakova, Sofia Egorova, Olga Kaspler, and Kira Vakhrusheva.

Finally, I thank my parents Nina and Guriy and my grandmother Sofia for always believing in me and supporting me in everything I do.

Chapter 1

Introduction

1.1 What this dissertation is about

The goal of this dissertation is to analyse the structure of the C-T (complementizer-tense) system in English and Russian, which will help me solve such important grammatical problems as Sequence of Tense in English and Russian and the peculiarities of Russian long-distance extraction. There are several ways in which Russian grammar is different from that of languages like English or Dutch. In what follows I will uncover general properties of language that underlie these seemingly diverse facts. The empirical phenomena I will concentrate on are Sequence of Tense and long-distance movement in Russian as opposed to languages like English. I will show that differences between the two (types of) languages in these domains are manifestations of the ways their respective C-T systems are organized. I will propose a theory of the structure of the C-T domains for English and for Russian, and I will go over some of the consequences that it has for the grammar of these languages.

I start this introductory chapter by outlining the set of data that I will be dealing with in this thesis (section 1.2). In the following section 1.3, I summarize the proposal that I am going to defend. Section 1.4 is dedicated to a brief overview of some aspects of Russian grammar that I will refer to in my discussion of this language. The final section 1.5 contains an overview of the dissertation.

1.2 Outline of the data

In this dissertation I explore cross-linguistic contrasts in the sphere of Sequence of Tense (SOT). Languages such as English or Dutch are said to display Sequence of Tense (SOT). Thus, an English sentence like (1), with a past tensed complement embedded under a past matrix verb, can have

two interpretations (1a) and (1b), which have been termed a ‘simultaneous’ and ‘past-shifted’ interpretations of the embedded past:

- (1) John thought that Mary was pregnant.
 - a. John thought that Mary was pregnant at the time of his thinking.
 - b. John thought that Mary had been pregnant at some previous time.

(2) demonstrates that present cannot normally be used to get a simultaneous interpretation when embedded under past:

- (2) # A year ago, John thought that Mary is pregnant.

In contrast, languages that are considered “non-SOT” languages use present and not past in a past-embedded complement clause to encode simultaneity between the events denoted by the matrix and the embedded verb. Russian is one of these languages, as shown in (3):

- (3) God nazad Ivan dumal, čto Maša boleet.
 Year ago Ivan thought that Masha ails
 ‘A year ago Ivan thought that Masha was ill (at the time of thinking)’

If past-under-past is used, the resulting reading of the embedded tense is normally past-shifted; a simultaneous reading is also possible but its occurrence is limited:

- (4) Ivan dumal, čto Maša bolela.
 Ivan thought that Masha ailed
 ‘Ivan thought that Masha had been ill (at some previous time) /
 (?)was ill (at the time of thinking)’

There are two issues connected with these facts: first, how do we derive the interpretations of embedded tenses, in particular, the simultaneous reading of the past tense in SOT languages? Second, how are non-SOT languages different from SOT languages and why do tenses behave differently in the two groups of languages? In this dissertation (Chapters 3 and 4), I will propose a theory of tense interpretation that answers both of these questions.

Another intriguing set of Russian data that I turn to in this dissertation is a well-known puzzle of Russian syntax – the limited character of long-distance wh-movement. The case of complement clauses is the most striking one. In English, long-distance extraction out of complement clauses is permitted:

- (5) Who do you think that Mary loves?

However, the Russian counterpart of (5) (extraction of the *wh*-word *kogo* ‘who-ACC’ across the clause boundary between the matrix clause and the embedded indicative complement clause with the complementizer *čto*) is degraded:

- (6) ?* *Kogo ty dumaeš, čto Maša ljubit?*
 whom you think that Masha loves
 ‘Who do you think that Masha loves?’

Judgements on (6) vary and range from “possible but only in very colloquial speech” to “completely ungrammatical”, but in any case (6) has to be marked as degraded. Interestingly, only indicative clauses behave in this way. Extraction out of a subjunctive complement clause embedded under a volition verb like *xotet’* ‘want’ is allowed:

- (7) *Kogo ty xočeš’, čtoby Maša poljubila?*
 whom you want that-SUBJ Masha love
 ‘Who do you want Masha to love?’

The data indicate that indicative complement clauses in Russian, as opposed to English, block extraction of *wh*-elements. This variation is left unexplained by the accepted views on locality of movement, expressed in theories of Subjacency, Barriers and subsequent theories. While some researchers (Bailyn, 1995a, 1995b; Stepanov, 2001; Szczegielniak, 1999, and others) have attempted to offer an explanation, no effective solution has been proposed. I will discuss this problem and present my own account of the data in Chapter 5.

The two sets of data that I have mentioned above seem to involve different grammatical domains: interpretation of tenses and *wh*-movement. However, what unites them is the syntactic configuration involved: (indicative) complement clauses, introduced by the complementizer *that* in English and by the complementizer *čto* “that” in Russian. In this dissertation I will present a theory that will derive the properties of English and Russian grammar discussed above from the clause structure of these languages.

1.3 The proposal

The proposal that I am going to defend is based on Phase Theory developed in (Chomsky, 2001, 2004, 2005). It introduces the notion of phases – “chunks” of syntactic structure that mark “stages” of the sentence derivation. This notion is especially relevant for the locality of syntactic relations, as Chomsky suggests that only the edge of the phase (phase head and its specifier, but not its complement) is visible for subsequent syntactic processes after the phase is built. Chomsky’s phase heads are C and *v*, and he

argues at length that T is not a phase, pointing to a close relation between C and T: without C, T is “defective”. This discussion will be at the core of my proposal. As for the actual character of the connection between C and T, I make use of the approach in (Pesetsky & Torrego, 2001, 2004a, 2004b) who propose that C carries an uninterpretable Tense feature.

The proposal contains the following claims:

1. The major claim I make about the grammatical structure of Russian is that T is a phase head in this language, and not C as in English. This accounts for the differences in the grammar of the two languages that I will explore.
2. I propose a theory of Sequence of Tense based on binding relations between tense morphemes. This approach has a history in the theory of tense interpretation, starting with (Partee, 1973) who suggested important parallels between tenses and pronominal elements. I view binding relations as based on Agree chains between T(ense) heads, following (Pesetsky & Torrego, 2004b) and (Reuland, 2001, 2005). It follows that the interpretation of tenses depends on the syntax of T.
3. The difference in the interpretation of embedded tenses in English and Russian complements comes from syntactic limitations on dependency relations between T-heads: these relations are possible in English but not in Russian. This is due to the fact that the Agree chain has to proceed via the embedded C head, which has different status and therefore different properties in Russian and English. In English, C bears an instance of T(ense) feature, which provides a link between the embedding and embedded T heads; in Russian, C does not bear T(ense). This difference is a corollary of the fact that C is not a phase head in Russian.
4. Long-distance *wh*-movement data in Russian are another consequence of the same crosslinguistic difference in phase structure: movement of the *wh*-phrase cannot pass the phase boundary in Russian. Although T is the phase head in Russian, C and T still form a unity, and CP is still the phase edge. However, as T is the phase head and attracts movement, moved elements cannot reach the phase edge position at Spec,CP.

1.4 Grammatical properties of Russian

This dissertation compares some aspects of Russian grammar with the grammatical properties of languages such as English and Dutch. In this section, I will summarize the properties of the grammar of Russian which

are relevant for the subsequent discussion: the tempofural/aspectual system of Russian, the structure of complement clauses, and the properties of wh-movement.

1.4.1 Temporal and aspectual system of Russian

The two major categories of the Russian verb are tense and aspect, which are very closely related. In this thesis I abstract away from the question of whether Russian displays a separate Aspect projection, since this is not important for my model. However, some discussion of the Russian aspect is necessary to understand the interpretation of tenses in this language. I will examine the two categories in turn, describing briefly their meaning and use. I will also summarize the formation and use of the Russian subjunctive.

Tense and aspect

Russian has two aspects, imperfective (IMPF) and perfective (PF). The division into the two aspects covers all verb forms (although several verb forms are ambiguous between the two aspects); in other words, every verb form in Russian is characterized as either perfective or imperfective. As for tense, according to traditional Russian grammars, Russian displays a simple tense system consisting of three tenses: past, present, and future. The paradigm of the Russian tense-aspect forms for the verb *delat'* (IMPF), *sdelat'* (PF) “do” in the indicative is presented in (8).

	Past (fem.sg.)	Present (1p.sg.)	Future (1p.sg.)
(8) IMPF	delá-l-a	deláj-u	budu delat'
PF	sdelá-l-a	—	sdeláj-u

As can be seen, the paradigm is incomplete. What is missing is the present perfective form. The verbal form *sdelaju*, which we would expect to find in the present perfective cell of the table, turns out to occupy the future perfective slot. Morphologically, it is a present form, displaying the same morphology as the corresponding imperfective form *delaju*. However, it cannot have a present interpretation; it is always interpreted as future. As for the imperfective future, it is the only complex tense form in the paradigm: it consists of the future form of the verb *byt'* ‘be’ plus the imperfective infinitive. Thus, it falls out of the general picture. These observations have led grammarians to argue that the Russian tense system has only one underlying opposition in its core: past vs. non-past, the latter comprising present and future. This view was famously proposed by (Jakobson, 1932), who suggests that the Russian tense system rests on a binary opposition +/- PAST, with the present/future tense receiving a negative characterization. This claim has been taken up in recent work, such as (Borik, 2002) and (Kusumoto, 1999).

While it is clear that tense is responsible for locating an eventuality on the time axis, the semantics of the aspectual opposition is notoriously difficult to define (for the discussion of aspect within the Russian linguistic tradition see e.g. (Bondarko, 1983, 1987, 1990; Maslov, 1984, 1985, a.o.)). In terms of markedness, perfective aspect is taken to be the marked member of the opposition, defined positively, and imperfective is unmarked and is often defined negatively as “non-perfective”. Putting it very informally, the perfective aspect represents the situation “as a whole” or as a clearly limited interval and looks at it “from outside” (Comrie, 1976). In contrast, the imperfective aspect views the situation “from within” and does not define its limits. It is possible to say that the meaning of imperfective overlaps with, but is on no account limited to, progressive semantics. The aspectual semantics is thus closely connected to the issues of telicity and point of view (for an extensive discussion, see (Borik, 2002)). Here are a couple of examples to demonstrate the semantic nuances of aspectual usage:

- (9) V prošlom godu Ivan stroil dom.
 in last year Ivan built-IMPF house
 ‘Last year Ivan was building/built a house’
- (10) V prošlom godu Ivan postroil dom.
 in last year Ivan built-PF house
 ‘Last year Ivan built a house (completely)’

Sentence (9), which contains an imperfective verb, does not give the information on whether or not Ivan actually finished the house he was building. In contrast, changing the verb form to perfective implies that the building of the house was completed last year. Still, this is not the end of the story. Sentence (11) demonstrates another use of imperfective, the so-called “statement of fact” use:

- (11) Ja uže videla etot fil'm.
 I already saw-IMPF this film
 ‘I have already seen this film’

Sentence (11) means exactly the same as its English translation: it states that the event described by the verb has taken place. In contrast to (9), the event here is viewed as completed. The list of various uses of the two aspects can go on; however, only the core uses demonstrated in (9) and (10) will be important for my discussion. In sentences with multiple predicates, two or more perfective predicates normally denote non-overlapping (temporally ordered) eventualities, while to denote a temporal overlap, an imperfective form has to be used:

- (12) Kogda Ivan prišel domoj, Maša prigotovila obed.
 when Ivan came-PF home, Masha cooked-PF dinner

‘When Ivan came home, Masha cooked dinner’ (in the sense, ‘after Ivan came home. . .’)

- (13) Kogda Ivan prišel domoj, Maša gotovila obed.
 when Ivan came-PF home, Masha cooked-IMPF dinner
 ‘When Ivan came home, Masha was cooking dinner’

Imperfective forms do not necessarily encode overlap, though; they can also be used in sentences where non-overlapping eventualities are reported. This is demonstrated by the (literal) translation of the English simultaneous past examples. Using past imperfective in the complement clauses gives both a past-shifted, i.e. non-overlapping, and possibly a simultaneous, i.e. overlapping reading:

- (14) John thought that Mary was ill.
 (15) Ivan dumal, što Maša bolela.
 Ivan thought that Masha ailed-IMPF
 ‘Ivan thought that Masha had been ill/(?)was ill’

On the other hand, using past perfective in a complement clauses will always trigger a past-shifted reading:

- (16) Ivan dumal, što Maša prigotovila obed.
 Ivan thought that Masha cooked-PF dinner
 ‘Ivan thought that Masha had cooked dinner.’

In view of this difference, I will always use imperfective past forms in complement clauses in my discussion of the SOT effects in Russian, because only imperfectives will provide the right context for an ambiguity between simultaneous and past-shifted readings. Establishing the exact mechanism that forces a past-shifted reading in case of perfective is outside the scope of this thesis, which concentrates on temporal relations and not on the aspectual meanings. I refer the reader to the literature dealing specifically with aspect, such as (Borik, 2002) and the references cited there, for approaches to Russian aspects that effectively derive temporal interpretations of perfective and imperfective verbs in cases like (12)-(13), and assume that they can be extended to cover the contrast in (15)-(16).

Subjunctive

The Russian verbal paradigm includes only one subjunctive form, which is formed by combining the past tense form of the verb, which is inflected for number and gender, with the uninflected particle *by*. There can be several positions for the particle in the clause; it acts as an enclitic to the verb or some other word in the sentence, so it cannot be clause-initial. In embedded clauses, it is usually attached to the (overt) complementizer or

wh-word. The past tense form used in Russian subjunctives does not carry any meaning of pastness, which is explained by the fact that historically the subjunctive form did not develop from the past form; rather, they both have a common source.¹ The subjunctive can have any temporal meaning; in contrast to English, there is no difference between past and present/future subjunctive. The use of subjunctive is illustrated in examples (17)-(18).

- (17) Esli by včera byla xorošaja pogoda, my pošli by
 If SUBJ yesterday be good weather we go SUBJ
 guljat'.
 to-walk
 'If the weather had been good yesterday, we would have gone for a walk.'
- (18) Esli by zavtra byla xorošaja pogoda, my pošli by
 If SUBJ tomorrow be good weather we go SUBJ
 guljat'.
 to-walk
 'If the weather was good tomorrow, we would go for a walk.'

The subjunctive is used in conditionals, as well as in root clauses with an irrealis meaning. In complement clauses it is required under volitional matrix verbs.

1.4.2 Complement clauses in Russian

This dissertation centers around Russian complement clauses. Before the discussion starts, however, a brief description of how Russian complement clauses are structured is in order.

Finite indicative complement clauses in Russian are normally introduced by the complementizer *čto* 'that':

- (19) Ivan znaet, čto Maša ljubit Petra.
 Ivan knows that Masha loves Petr

The form *čto* also functions as the wh-word 'what':

- (20) Čto ty delaeš?
 what you do
 'What are you doing?'

¹In Old Russian, the subjunctive mood form consisted of the aorist form of the auxiliary verb *byti* 'be' and the past participle, which contained the affix *l* and agreed with the subject in person and number. Later, the auxiliary lost its inflection, and only one of its forms continued to be used: *by*, the 2nd or 3rd person singular of the aorist, which turned into the subjunctive particle. The participle thus became the inflected form. Past tense, on the other hand, developed from the perfect, which was originally also formed of the past participle combined with an auxiliary. The auxiliary was later lost completely, and the participle took on the role of an inflected past tense form.

Complement clauses in the subjunctive mood have a different form of the complementizer – *čtoby* (with a variant *čtob*). Subjunctive is required in complements of volitional verbs like *xotet'* 'want', verbs of request like *prosit'* 'ask' and some others. It is also possible, but not obligatory, in complements of verbs of thinking (especially with negation or with matrix verbs expressing doubt):

- (21) a. Ivan xočet, čtoby Maša ljubila Petra.
Ivan wants that-SUBJ Masha love Petr
'Ivan wants Masha to love Petr'
- b. Nel'zja skazat', čtoby etot professor byl geniem.
Impossible to-say that-SUBJ this professor be genius
'One can't say that this professor is/was a genius.' (Brecht, 1977, p. 34, (5b))

There is sufficient evidence for analysing *čtoby* as the complementizer *čto* which is glued together with the subjunctive particle *by*, which is part of the verbal subjunctive form (here *ljubila by*). First, as can be seen from the previous section, *by* can be used independently of *čto*: it is a clitic-like element that is always adjacent to the complementizer when subjunctive occurs in an embedded clause. Apart from the subjunctive form, *by* can also be combined with infinitive to express an irrealis meaning, and with some adverbials:

- (22) a. Vypit' by sejčas čaju!
to-drink BY now tea
'I wish I could drink some tea now!'
- b. Pëtr vrode (by) ljubit Mašu.
Petr seemingly (BY) loves Mary
'Peter seems to love Mary'

Importantly, as noted in (Brecht, 1977), if the embedded clause headed by *čtoby* contains more than one predicate, the particle *by* can be repeated with the second (and subsequent) verb(s) without the *čto* part, as in (23); this shows that *by* is indeed part of the verbal form and not of the complementizer. Referring to *čtoby* as a complementizer is thus only done for convenience: in fact, the complementizer is *čto* and *by* is a clitic attached to it.

- (23) a. Ne xoču ja, čtob ty pritvorjalasja i k ulybke
Not want I that-SUBJ you pretend and to smile
sebja prinuždala by.
yourself force SUBJ
'I don't want you to pretend and force yourself to smile'

- b. Ty velel, čtoby ja uexal v Minsk odin, a Vasja
 you ordered that-SUBJ I go to Minsk alone and Vasja
 by ostalsja s toboj?
 SUBJ remain with you
 ‘Did you order that I leave for Minsk alone and Vasja remain
 with you?’ (Brecht, 1977, p. 35-36, (6d-e))

Infinitival clauses are usually not introduced by any complementizer, with the exception of infinitival clauses of purpose, which allow, and sometimes require, the complementizer *čtoby*:

- (24) a. Ja xoču kupit’ xleba.
 I want to-buy bread.
 b. Ja pošla v magazin, (čtoby) kupit’ xleba.
 I went to shop (that-BY) to-buy bread
 ‘I went to the shop (in order) to buy bread’
 c. *(Čtoby) kupit’ xleba, nužno pojti v magazin.
 that-BY to-buy bread necessary to-go to shop
 ‘In order to buy bread, one needs to go to a shop’

An interesting property of the Russian complement clauses is that they are regularly combined with a so-called pronominal correlative in the main clause; the correlative is a form of the pronoun *to*, which is the neuter singular of the demonstrative pronoun *tot* ‘that’. The correlative is selected by the matrix verb, as a complement or inside a PP, and the complement clause is in turn attached to the correlative. The term “pronominal correlative” stems from (Comrie, 1971), however, it might be more accurate to call this element a “pronominal substitute”. Often, instead of embedding a subordinate clause into a certain syntactic position, it is possible to put the corresponding form of *to* into this position, and then attach the clause to it.

- (25) Ja pomnju (o tom), čto Petr ljubit Mašu.
 I remember (about it) that Petr loves Masha.

The correlative usually occurs in positions where it carries oblique case or is embedded into a PP. It is least frequently found in accusative case, unless it carries a contrastive stress (Švedova, 1982). In that case, it is used to put logical stress on the complement, as in (26):

- (26) Ja znaju **to**, čto Ivan pobil Petra.
 I know it that Ivan beat Petr
 ‘I know that Ivan has beaten Petr (and this is a thing I am really sure about)’

The correlative can also be modified by focus particles like *tol’ko* ‘only’, as in 27. The particle then refers to the whole embedded clause.

- (27) Ja znaju tol'ko to, čto ja ničego ne znaju.
 I know only it that I nothing not know
 'The only thing I know is that I know nothing'

To my knowledge, no special analysis of the pronominal correlative construction has been proposed in the literature, with the exception of (Stepanov, 2001). For Stepanov, all "complex NPs containing finite clausal complements actually involve adjunction of the clause" (Stepanov, 2001, p. 149). Consequently, he analyses the correlative as an NP and the clause as adjoined to it. I will not hold to Stepanov's general statement about complex NPs, but it seems reasonable to analyse the clauses in sentences like (25)-(27) as adjoined to the correlative. An alternative would be to view them as genuine complements; however, semantically the clause does not appear to be selected by the correlative NP, rather, it simply explicates the content of the correlative.

1.4.3 Wh-movement and wh-questions in Russian

Russian displays wh-fronting in root questions:

- (28) Kogo Maša ljubit? - Maša ljubit Ivana.
 who Masha loves - Masha loves Ivan
 'Who does Masha love? - Masha loves Ivan.'

This fronting is obligatory except for some specific cases.² Like many other Slavic languages, Russian also displays multiple wh-fronting:

- (29) Kto kogo pobil?
 who whom beat
 'Who beat whom?'

In multiple wh-questions, Russian does not display any superiority effects among the fronted wh-phrases (although their order is not arbitrary, being defined by information structure). (Stepanov, 1998) and, following him, (Bošković, 2002) argue that in Russian, all wh-phrases are fronted as focused phrases, so no obligatory wh-feature driven movement takes place. However, there are at least two problems with this approach:

1. (Stjepanović, 1999) argues that in Serbo-Croatian, and some other languages for which the focus-driven account is proposed, contrastively focused phrases are obligatorily fronted. This does not seem to be the case for Russian. So there is no reason why wh-phrases must be obligatorily fronted, which normally happens in Russian.

²For example, in questions with a pronominal subject the wh-word often follows the subject: Ty gde? (You where?) = where are you?

2. (Bošković, 2001, 2002) claims that overt wh-movement to Spec,CP (not to the focus position) excludes single-pair answers to multiple wh-questions. Thus, the English question “Who bought what?” is understood as relating possible buyers to possible purchases, not as a simple question with two unknowns; it can be answered with a series of buyer-purchase pairs: “John bought a book, Mary bought a pen . . .”, but not with one single pair: “John bought a book”. Boscovic uses this test to show whether a language has “real” wh-movement which checks a wh-feature. (Stepanov, 1998) claims that single-pair answers are possible for Russian multiple Wh questions. However, the judgements given in his paper are not shared by the speakers that I consulted. Rather, it seems that only pair-list answers are possible for questions like that in (30):

- (30) Kto čto kupil? – Ivan kupil knigu, a Maša – žurnal. –
 Who what bought? – Ivan bought book, and Masha – magazine –
 #Ivan kupil knigu.
 #Ivan bought book
 ‘Who bought what? – Ivan bought a book and Masha bought a magazine. – #Ivan bought a book’.

So if we take Bošković’s generalization to be right, Russian patterns with languages which have ‘real’ wh-movement. Without going deeply into the discussion, I will adopt the view that wh-movement in Russian proceeds by means of the usual wh-feature checking mechanism.

1.5 Structure of the dissertation

The rest of this dissertation is organized as follows. Chapter 2 presents the general theoretical background of my research. The first part of the chapter (section 2.1) deals with the notion of the C-T (complementizer-tense) system. In section 2.1.1, I trace the development of this notion at various stages of generative grammar. The following section (2.1.2) contains an overview of Phase Theory (Chomsky, 2005, 2006). Phase Theory stresses the close connection between C and T: in this framework, CP is a phase, and TP is dependent on it, and defective without it. In section 2.1.3, I analyse the work of Pesetsky & Torrego (2001, 2004a,b), who propose that C in English bears an instance of T feature and develop a theory of Agree as feature sharing. Their proposal is crucial for my discussion, so I will dwell on it in some detail.

The second part of Chapter 2 (section 2.2) is dedicated to Binding Theory, which is crucial for the discussion since my proposal regarding the interpretation of embedded tenses is based on the notion of tenses as referential expressions which can receive a bound interpretation. In section 2.2.1,

I introduce the basic principles of the “classical” binding theory (Chomsky, 1981), which defines binding in terms of c-command and coindexation and formulates the three syntactic binding principles A, B and C to rule out ungrammatical instances of binding. In section 2.2.3, I turn to the theory of binding developed by (Reinhart, 1983, et seq.), (Reinhart & Reuland, 1993), who argue that the interpretation of pronouns and anaphors is governed by distinct processes of binding, coreference (or coindexation) and reflexive-marking of predicates. In section 2.2.4, I analyse the model of binding of SE (simplex) anaphors proposed in (Reuland, 2001, 2005), which is based on Pesetsky & Torrego’s Agree mechanism.

In Chapter 3, I present the theory of Sequence of Tense (SOT). First, in section 3.1 I make an overview of the data that are in the focus of SOT studies: the interpretation of tenses in complement clauses, especially the readings of past tense embedded under past; the interpretation of tenses in relative and adjunct clauses. I also demonstrate the crosslinguistic differences in the sphere of SOT, paying special attention to SOT facts in Russian. In section 3.2 I present the accounts of SOT. Since the literature on the subject is extensive, I limit myself to several theories that I will rely on in developing my own proposal (Enç (1987, 2004), Abusch (1988, et seq.), Schlenker (1999, 2003, 2004), Kusumoto (1999) and some others).

I present my account of SOT in English (and other SOT languages such as Dutch) and Russian in Chapter 4. In section 4.1 I address the question of how the so-called “simultaneous” reading of past tense embedded under past in complement clauses is derived, and argue that some cases of this reading are better analysed as irrealis uses of the past morpheme. I present evidence that shows that the simultaneous reading of past is also available in a non-SOT language like Russian, however it can be proved that this reading is derived differently from the English simultaneous past. In section 4.2 I formulate my account of SOT in SOT languages like English or Dutch, based on an Agree relation between T heads, and explain why Russian behaves differently from them. Some remaining issues are addressed in section 4.3.

In Chapter 5, I turn my attention to a different set of problems, namely, the restrictions on long-distance extraction in Russian complement clauses. Section 5.1 contains general discussion of locality; I discuss the theory of successive-cyclic movement and island constraints, as well as several approaches to locality of movement. In section 5.2 I introduce the Russian locality facts, which have so far resisted explanation. I also present several accounts of long-distance extraction in Russian and highlight the problems that they raise. In section 5.3, I develop my account of long-distance movement in Russian and summarize the properties of the Russian phase structure which are at the core of both the interpretation of tenses and the movement facts in Russian complement clauses.

Chapter 6 contains concluding remarks and some perspectives for future research.

Chapter 2

General syntactic background

In this chapter, I will outline the general background necessary for my research. In the first part of the chapter (section 2.1), I will talk about the modern views on the C-T system. After discussing clause structure in general (section 2.1.1) I will focus on Chomsky's phase theory, which builds on the idea of a close relationship between the C and T heads (section 2.1.2); then I will summarize the theory of Pesetsky and Torrego, who provide a concrete technical implementation of this idea (section 2.1.3). This part of the chapter is summarized in section 2.1.4.

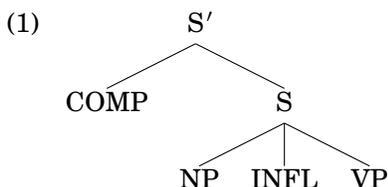
In the second part of the chapter (section 2.2), I will present Binding Theory, which I will use in my analysis of temporal relations. In section 2.2.1, I will introduce the key notions of the "classical" binding theory of Chomsky (1981). In section 2.2.3, I will give an overview of a contemporary take on binding theory by Reinhart (1983 et seq.), Reinhart & Reuland (1993). Finally, in section 2.2.4, I will present the approach to binding of anaphors developed in (Reuland, 2001, 2005). I summarize this part of the chapter in section 2.2.5.

2.1 Clause structure and the architecture of the C-T system

2.1.1 The C-T system

The present dissertation deals with the structure of the C-T system, in particular, with the relation between the C (Complementizer) and T (Tense) heads and its crosslinguistic variation, so first it is necessary to establish what is understood under the notion "C-T system". In this section I will talk briefly about the development of this concept in generative grammar.

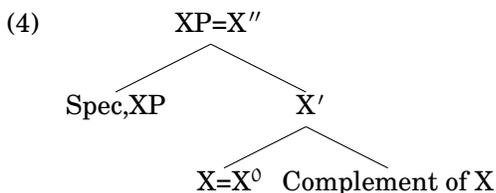
The notion of the C-T *system*, that is, the idea that the C and T (or I, Inflection) layers of syntactic structure form a whole, can be traced back to the early work on the Government and Binding (GB) theory. The two categories – C and T (I) – were born from one category S (Sentence), which was the highest category in the sentence structure in the earliest versions of generative grammar: rewriting rules transformed S into NP (the subject of the sentence) and VP, and so forth. As the theory evolved, the category S became more complex. In the synthesis of the work on the topic presented in (Chomsky, 1981), a higher category was added to the sentence structure, necessary primarily for embedded clauses and questions – S', which hosted complementizers, such as *that*, *for*, etc, as well as *wh*-words. (Chomsky, 1981) also introduced a functional category INFL (Inflection) which was characterized by values [+Tense] or [-Tense] and served to determine whether the clause was finite or non-finite (for example, infinitival). This category was separate from the VP. Thus, a clause now looked the following way:



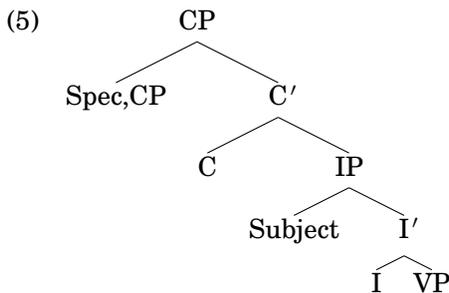
(Chomsky, 1981) stresses a connection between finiteness and the element occupying COMP. For example, in (2) INFL is [+Tense], and the complementizer *that* is chosen; in (3), INFL is [-Tense] and the complementizer is *for*; the corresponding structures are given in the (b) examples.

- (2) a. John knows that Bill saw Mary
 b. John [_{VP} knows [_{S'} that [_S Bill [+Tense] [_{VP} see Mary]]]]
- (3) a. the students prefer for Bill to visit Paris
 b. the students [_{VP} prefer [_{S'} for [_S Bill [-Tense] [_{VP} visit Paris]]]]

At the same time, the similarities between the lexical categories N, V, A and P were captured by the X'-theory, which provided a uniform internal structure for all these categories, the X'-structure: a lexical head X (X⁰) projected to form the intermediate-level projection X' and the maximal projection XP (X'')



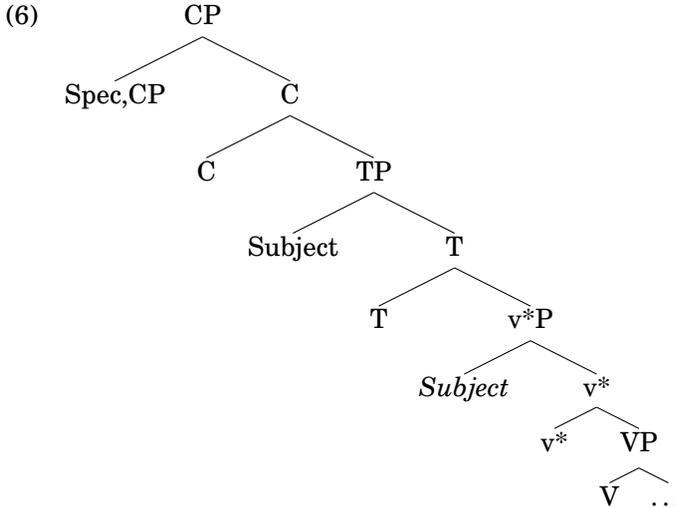
Keeping the categories S' and S in the clause structure meant that the functional categories COMP and INFL were dramatically different from the lexical categories because they did not head their own projections. As a consequence, functional categories were exocentric, while the rest of the categories were endocentric. In later versions of the theory, for example, in (Chomsky, 1986), this discrepancy was eliminated: S and S' were no longer present in the structure, and the X'-theory was expanded to cover the clausal categories as well. This meant that both C and I projected: S' turned into CP and S turned into IP. Now, CP became the highest category, it dominated IP, so that the upper part of the clause received the now familiar aspect:



After CP and IP became separated, the clause was viewed as consisting of three layers: the lexical layer, responsible for the assignment of theta roles – the VP; the inflection layer – the IP, associated with the licensing of argumental features such as case and agreement; and the complementizer layer, the CP, which was present not only in clauses headed by a complementizer, but in all other clauses as well, and hosted topicalized elements and different operators such as *wh*-words (Rizzi, 1997). Each of these layers was necessarily multifunctional, carrying several grammatical features, and in the late eighties they in turn were split, so that virtually every feature was given a separate projection; at the time, this provided a more principled way of computation compared to dealing with feature bundles. Now both CP and IP became split: starting from (Pollock, 1989), IP was divided into a tense projection (TP), agreement projections for Subject and Object (AgrSP, AgrOP), an aspectual projection (AspP), and others (including adverbial projections proposed by (Cinque, 1999)); CP, the “left periphery” of the clause, was taken to consist of a force projection, a finiteness projection, and possibly topic and focus projections (Rizzi, 1997). VP was also granted several layers for multi-argument verbs (Larson, 1988, a.o.).

Later, in (Chomsky, 1995) as well as in the work of many other authors, the division of IP into TP and Agreement projections was rejected. Instead, a single T node was postulated, containing both temporal features and agreement (ϕ) features for the subject. As for AgrOP, its role was taken over by the part of the split VP projection termed *v*P. *v*P (little *v*

phrase) is located higher than VP, and the two verbal projections provide positions for the external and internal arguments, respectively: the subject of the sentence originates in the Spec,vP and can later move to Spec,TP; the internal arguments occupy the positions of specifier and complement of V (Chomsky, 1995). Apart from this modification, I will not be concerned with the internal structure of TP and CP in this dissertation, so I will be assuming a simplified clause structure given in (6).¹



2.1.2 Phase theory

The version of generative theory that I am adopting in this thesis is Chomsky's Phase theory, first formulated in (Chomsky, 1999) and elaborated in (Chomsky, 2005, 2006). In what follows I will present an overview of Phase theory in (Chomsky, 2005, 2006), highlighting the points that will be of special importance for my own research.

In the Phase theory framework, two basic syntactic operations are postulated: Merge and Agree. Merge involves putting together two lexical items X and Y, a lexical item X with a (complex) syntactic object Y, or two syntactic objects, to form a new syntactic object {X,Y}. Two varieties of Merge are distinguished: External Merge, where X is not part of Y, and Internal Merge (IM), that is, the familiar Move operation, where X is part of Y and gets displaced and attached to Y. Merge is regulated by the No-Tampering Condition (NTC), which prevents it from changing the participating elements in any way; thus, Merge can only attach one object to the edge of the other – it cannot look inside, and IM has to leave behind a copy

¹I will follow (Chomsky, 2005) in using the notation v* for all verbs except for unaccusative and passive, and I will occasionally use v without a star when quoting the literature, or when the argument structure of the verb is not specified.

of X within Y. The motivation for this becomes clear if we consider the set-theoretic notation of sentence derivation. Each syntactic object carries a certain label, and when two objects X and Y are combined by Merge, the resulting new object receives the label of either X or Y. The result then looks like $\{X, \{X, Y\}\}$. Since the information about the internal structure is preserved, Merge is forbidden to tamper with the objects that it combines.

Another basic syntactic operation is Agree; both IM (involving displacement of a syntactic object) and Agree (matching not based on displacement) rely on a probe-goal relation. A probe is an object with unvalued features that scans its domain for a suitable goal containing valued instances of the corresponding features. Once the goal is found, an Agree relation is established between it and the probe, so that the goal values the unvalued features of the probe. (Chomsky, 1999) proposes that not only the probe but also the goal has to be *active* for Agree to take place, that is, it has to carry uninterpretable features. This view limits Agree to operations involving Case-checking of DPs: (Chomsky, 2005) specifies that in the course of the Agree operation the probe also values the structural Case feature on the goal. Thus, the Agree relation is always mutual – both the probe and the goal profit from it. Operations involving A'-positions are now exempt from Agree.

The notion of phases, central to the theory, is introduced to explain the phenomenon of successive cyclicity in language. (Chomsky, 2005) proposes that at certain stages of the derivation, the syntactic object already constructed is handed over to the phonological and semantic component by means of Transfer operations. The phonological component then maps the structure to the sensory-motor (SM) interface, and the semantic component – to the conceptual-intensional (C-I) interface. Syntactic objects to which Transfer applies are called phases. Furthermore, Chomsky argues that along with Transfer, all other operations except for External Merge should also apply at phase level. Consequently, only phase heads can be probes for Agree and IM, so movement can only proceed to the phase edge.

The next question is, of course, what syntactic objects qualify as phases. By their intrinsic nature, phases should mark some logical intermediate stages in the course of the derivation at which it becomes relatively complete. (Chomsky, 2005, 2006) proposes CP and v^*P for this role. CP here is meant to be “shorthand for the region that (Rizzi, 1997) calls the “left periphery”, possibly involving feature spread from fewer functional heads (maybe only one)” (Chomsky, 2005, p. 9). Chomsky denotes by v^* “the functional head associated with full argument structure, transitive and experiencer constructions, and is one of several choices for $v(\dots)$ ” (ibid.). It follows that the choice of v^*P is justified naturally because it represents a predicate complete with all its arguments that have received their theta-roles. In other words, v^*P is thematically complete. Since all syntactic operations apply at phase level, Chomsky concludes that CP must also necessarily be a phase, because A'-movement targets its edge. However,

along with CP, there appears to be another plausible candidate for phasehood: TP. The argument in favour of it being a phase is that it is the bearer of the temporal features and the ϕ -features involved in the Nominative-agreement system. However, Chomsky argues extensively against the phasehood of TP. His main argument is the following: the Tense feature and the ϕ -features on T are not there in the lexicon. Rather, they are inherited from C when C is merged to T. This is demonstrated by the fact that T only manifests these features when selected by C, otherwise it can only appear in a raising or ECM infinitival clause, but not in a fully-fledged clause.

(Chomsky, 2006) provides another argument proposed by Marc Richards (p.c.): “TP cannot be a phase, with operations of valuation and A-movement driven by properties of T. Suppose TP were a phase. Then its interior will be transferred by PIC [phase-impenetrability condition], but the head T will retain its valued uninterpretable features. The derivation will therefore crash at the next phase . . .”; “the uninterpretable features of C must be “inherited” by T. If they remain at C, the derivation will crash at the next phase. Note that TP cannot be saved by the same device: its features cannot be inherited by v^* , which already has features.” (p. 13) The general idea behind this argument is that C always has to be complete and cannot carry any uninterpretable features whatsoever; this is also the intuition behind the edge feature theory (see below).

Chomsky concludes that the phase-like properties of T are derivative of C. This idea is novel to the theory, and crucial for the proposal that I will be presenting. It implies that CP and TP are dependent on one another; together they form a unit that takes part in syntactic processes as a single whole. Thus, it is the C-T combination that attracts a DP goal (a subject), although it only raises to Spec,TP due to the fact that it is T that (derivatively) bears the relevant features. It should be emphasised that although CP gives away its features to TP, CP is still, informally speaking, “more important” than TP: metaphorically, CP is the “employer” which assigns certain jobs and responsibilities to TP, its “employee”. Because it dominates TP, CP receives the phase role.

Phase theory also has implications for the locality of movement, by virtue of the Phase Impenetrability Condition (PIC). The PIC specifies that once the information contained in a phase is transferred to the interfaces, it is “forgotten” by the computation system and cannot be accessed any more. Only the edge of the phase is still visible for subsequent syntactic operations. This makes the derivation strictly cyclic, and serves as an explanation for the successive-cyclic character of A'-movement. Obviously, all the elements that need to be moved across a phase boundary should be situated on the edge of the phase by the time the completed phase is merged further, so that they could be moved on into the higher phase later. Island constraints on movement are explained by introducing the idea of a search domain which is visible for each phase.

The nature of A'-movement remains to be specified because, as was noted above, in the Phases framework Agree is no longer responsible for any operations involving A'-positions, in particular, for A'-movement such as wh-movement. Thus, a different motivation is necessary. (Chomsky, 2005) claims that all A'-movement to the phase edge is free and not Agree-driven. Phases are endowed with *edge features*, which are responsible for free attraction of any available elements. Since all operations target phase edges, no other heads except for phase heads can get these features. What particular element moves depends on its interpretation in the derived position.

In developing my own proposal, I will largely adopt the Phase theory formulated in (Chomsky, 2005, 2006), with some modifications. The main one of them concerns A'-movement and the theory of edge features. (Slioussar, 2007) argues that it is problematic to view wh-movement as absolutely free and not dependent on Agree. She analyses Information Structure(IS)-related movement that underlies Russian word order permutations (scrambling) and comes to the conclusion that IS-related movement is indeed independent of the Agree relation and is driven by edge features only; however, she argues that wh-movement is still best analysed as feature-driven. I will stick to Slioussar's (2007) view in my discussion, and I will not maintain the view that Agree can only involve active goals. I will come back to successive-cyclic movement in Chapter 5.

2.1.3 Pesetsky and Torrego's theory of C-T dependency

The idea of an intrinsic connection between C and T is given a specific technical development by (Pesetsky & Torrego, 2001, 2004a, 2004b), who propose an implementation of Chomsky's insight that I will adopt in this dissertation. In short, their theory is based on the claim that a T(ense) feature is present on C. In this section I will summarize the arguments that led Pesetsky and Torrego to this conclusion, and the consequences that their approach entails. I will also pay attention to the theory of Agree proposed in (Pesetsky & Torrego, 2004b).

Pesetsky and Torrego 2001

In their 2001 paper, Pesetsky and Torrego (hereinafter PT) start out by enumerating several empirical phenomena that are quite mysterious for the current syntactic theory. The central problems for their research are numerous subject/non-subject asymmetries observed in English as well as many other languages. Below I give the relevant examples from (Pesetsky & Torrego, 2001) and discuss their solution to the problems. Before I do this, however, it is necessary to summarize the theoretical stance that PT take with respect to the syntactic operations they discuss.

PT (2001) stress that they adhere strictly to Chomsky's (Chomsky, 1999,

2000) claim that movement is never optional or unmotivated: it is always triggered by the presence of agreeing features on the attractor head and the attracted element. To this they add their own implementation of the distinction between Agree and Move. In case of Agree, an element X bearing an uninterpretable feature F establishes a connection with another element Y bearing an interpretable instance of the same feature. The uninterpretable feature is then deleted – which for PT does not mean that it disappears immediately; rather, it is “marked for deletion” but does not disappear until a later point in the derivation (possibly, until the completion of the current phase). As for Move, it happens if the uninterpretable feature responsible for establishing the Agree operation has the EPP property – the property of attracting material from Y to X. Hence, for PT the EPP feature is not a property of a head, but a property of a feature – thus it is not a feature on a par with others, but rather a subfeature. I will not adopt this view on EPP (or rather, this use of the term “EPP”); however, it is necessary to keep it in mind for the discussion of PT’s claims.

The idea that the operations Agree and Move only occur when they are strictly necessary is formulated as the Economy Condition (PT’s (6), p. 359): “A head H triggers the minimum number of operations necessary to satisfy the properties (including EPP) of its uninterpretable features”. This means that if certain features of H can be deleted by application of, say, either two or three different operations, the system will always opt for applying two operations and not three. Even if the result is the same, the Economy Condition dictates a more “sparing” choice.

The first empirical problem considered by PT is the so-called T-to-C asymmetry. In English root questions, the head T normally has to move to C (and gets spelled out as a form of the auxiliary *do*); however, this movement is prohibited if the *wh*-phrase moved to C is the subject. This is illustrated in (7) (PT’s (5)), where (7b) demonstrates the obligatoriness of T-to-C movement in case the object is *wh*-fronted, and (7c) demonstrates its impossibility if the fronted element is the object:

- (7) a. What did Mary buy?
 b. * What Mary bought?
 c. * Who did buy the book? [* unless *did* is focused]
 d. Who bought the book?

The first step in the argumentation is to establish why T has to move to C at all, given that all movement is necessarily triggered by some uninterpretable features. PT propose that this instance of movement is no exception: C bears an uninterpretable T feature (uT) with an EPP property. Consequently, C must agree with an element bearing an interpretable T feature (that is, with the T head) and attract it due to the EPP property of its uT feature.

Since (7d) does not exhibit T to C movement and is still grammatical, it is natural to propose that the uT on C is deleted by some other element

and not by T. PT claim that here, the *wh*-word moved to C plays this role. This is possible since the moved *wh*-phrase is a subject, hence it bears nominative case, which has a special property (PT's (8)): "Nominative case is uT on D". If this is so, additional movement of T to C is unnecessary. Moreover, it is, in fact, impossible if we consider the Economy Condition: movement of the subject *wh*-word can satisfy both the uninterpretable *wh*-feature on C (uWh) and the uT on C in only one step, so an additional movement step is rejected.

PT proceed to claim that not only C heads of matrix questions, but also other types of Cs bear uT – eventually, all finite C's are unified as having this feature. Based on this assumption, PT propose solutions for a number of further problems, such as the notorious *that*-trace effect first discussed in (Perlmutter, 1971). Sentence (8a) (PT's (3a)) demonstrates that when an object *wh*-word undergoes long-distance movement out of a complement clause, the complementizer *that* can be both pronounced and omitted. However, as (8b) (PT's (3b)) shows, if the *wh*-moved element is the subject, *that* is obligatorily omitted, or, in somewhat old-fashioned terms, it cannot be adjacent to the subject trace:

- (8) a. Who do you think (that) Sue met ___?
 b. Who do you think (*that) ___ met Sue?

PT note a parallel between the *that*-trace effect and the T-to-C asymmetry: "In both cases, subject *wh*-extraction prevents a word from occurring in C that is found there otherwise. In the case of the T-to-C movement asymmetry in English, the element barred from C with subject *wh*-movement is the tensed auxiliary verb, which we analyze as an instance of T that has moved to C. In the case of the *that*-trace effect, the element barred from C with subject *wh*-movement is the word *that*, which we are accustomed to thinking of as the complementizer itself, inserted as the sister of TP by Merge" (p. 371). According to PT, it follows that the root of the problem is the same as before: the embedded C bears uT which has to be deleted somehow, and movement of the subject across it influences this process. However, clearly there is no T-to-C movement in embedded clauses in Standard English. PT propose that the element which moves to C in order to check its uT feature is actually the word *that*: it is not a complementizer (which is null in this case) but an instance of T moved to C.

This is how the explanation works: just as in the previous case, the embedded C bears two uninterpretable features: uWh (to trigger movement of the *wh*-word into an intermediate "landing site" at the edge of the embedded clause) and uT. If the object *wh*-word is extracted, it satisfies the uWh feature on C; the uT feature on C is checked by T-to-C movement, which is represented by the appearance of *that*. However, if the moved element is not an object but a subject, thus bearing nominative case, both features can be simultaneously satisfied by it: the *wh*-feature deletes the

uWh, and nominative deletes uT. Again, additional movement of T into C becomes both superfluous and impossible.

What about cases when *that* is absent but the wh-element is not a subject and thus cannot satisfy uT on C? Since PT argue that not only C's bearing uWh, but all English C's bear uT, the question applies not only to instances of object wh-movement, but to all instances of embedded declaratives, for example in sentences like *Mary thinks Sue met John*. The answer is that uT on C can be satisfied not only by T itself but also by its Spec, i.e. the subject. Indeed, this must be possible since the nominative case of the subject is also an instance of uT. By PT's definition of closeness (which I will not go into for the sake of brevity), both T and its Spec are equally close to C, so it can choose freely which of the two will check its uT feature. In case the subject is chosen, it raises to Spec,CP. In fact, this option is available also in matrix clauses with an obligatorily fronted object wh-word. However, choosing subject movement instead of T-to-C movement has an effect on the interpretation – we get an exclamative instead of a question, as shown in (9) (PT's (42a) and (43b)):²

- (9) a. What did Mary buy?
 b. What a silly book Mary bought!

It should be noted that while (Pesetsky & Torrego, 2001) propose that *all* instances of finite C bear uT in English, they refrain from any definite claims about declarative matrix clauses. Mentioning the problem in notes (41) and (44), they say that they leave open the question whether matrix declarative clauses are CPs or TPs. Obviously, if they are TPs, the discussion is not applicable to them. If they are considered CPs (as, in fact, is commonly assumed), PT suggest that a certain mechanism might prevent uT on their C heads from being checked by movement of T in the shape of *that*, so that it can only be done by movement of the subject into CP. In particular, they think, “[i]t could be that *that* is available as a realization of T only when in the c-command domain of another occurrence of T (. . .)” (p. 411). However, obligatory movement of the subject into C might be incompatible with theories that argue for a rigid functional head hierarchy, like (Cinque, 1999). An alternative that PT do not discuss could be checking uT on C not by movement but by agreement with the T head or the subject (in PT's model, that would mean that in this case uT has no EPP property). This option is in fact considered available by PT, who argue that this is what happens in embedded interrogatives in Standard English, as shown in (10) (PT's (47a,b)):

- (10) a. Bill asked what Mary bought.

²What is left unclear in this formulation is how uT on the subject is supposed to check uT on C if both of them are not interpretable. In (Pesetsky & Torrego, 2004b), this issue is clarified: uT on the subject gets valued together with uT on the T head after agreeing with the finite verb *v*.

- b. * Bill asked what did Mary buy.

PT propose that C in embedded questions like (10a) bears uT without an EPP property, so it can be satisfied by agreeing with the subject DP or the T-head itself, without any movement. Interestingly, in Belfast English T-to-C movement is found in embedded questions (i.e. sentences like (10b) are grammatical), so an instance of parametric variation is proposed: uT on the embedded interrogative C does have the EPP property in Belfast English.

In this thesis I will mostly discuss the properties of embedded C's, so the question of matrix declaratives is not of central importance. However, in line with the argumentation behind the current phase theory, I will stick to the view that they are indeed CPs; it follows that their C head bears uT , and I will assume that it is satisfied by agreement.

PT's assumptions about embedded declarative clauses allow them to propose a solution for yet another puzzle of the English syntax – the so-called *that*-omission asymmetry, illustrated in (11) (PT's (4), taken from (Stowell, 1981)). While we can freely omit *that* when the embedded clause is in the complement position, *that* becomes obligatory if the clause occupies the position of the matrix subject:

- (11) a. Mary thinks [that Sue will buy the book].
 b. Mary thinks [Sue will buy the book].
 c. [That Sue will buy the book] was expected by everyone.
 d. * [Sue will buy a book] was expected by everyone.

From PT's perspective, the difference between the embedded clauses in (11c) and (11d) is that in the first one, uT on C triggers T-to-C movement spelled out as *that*, and in the second one, uT is interpreted via movement of the subject DP (*Sue*). PT capitalize on this contrast and propose that the different behaviour of the two clauses is caused by the difference in the feature composition of their C heads by the time the embedded clause is merged with the matrix. Recall that PT assume a possibility of delayed deletion for uninterpretable features: after an uninterpretable feature enters into an Agree relation, it is "marked for deletion" and can be finally deleted at some later stage. Importantly, though, this delay is not supposed to be obligatory, and the feature can be forced to completely delete by independent considerations. This is what happens in the case of (11d): the T feature on *Sue* must disappear, otherwise the subject might be attracted further from its position in Spec,CP, for example by the matrix T, forming a sentence like the ungrammatical (12) (PT's (60)):

- (12) * Sue was expected [_ will buy the book].

PT propose that the presence of the T feature on the DP is crucial for its ability to be attracted by a T-head. To generalize over this consideration,

they formulate the “Match Condition” (their (61)): “If a head H enters an Agree relation with a set of phrases K, each syntactic feature of H must be present on some member of K (not necessarily with the same value, including the value for EPP)”. This condition is also necessary to distinguish between (11c) and (11d). The embedded CP can only be attracted to the matrix TP if it has all the features of the T head, including the T features. However, the embedded CP in (11d) does not qualify because there is no instance of the T feature, the T feature of the subject being already deleted. In (11c), on the contrary, the embedded CP is headed by *that*, which is the T head itself that has moved to the CP and which obviously bears an interpretable T feature. It is therefore allowed to be merged to the matrix TP, and the derivation converges.

Concluding the discussion of the *that*-omission asymmetry, PT point out to an interesting cross-linguistic contrast. They quote Polish examples demonstrating that in Polish, embedded clauses cannot be subjects even when they are headed by the complementizer *że*. However, the sentence can be rescued by adding the element *to* to the complementizer, as is shown in (13) (PT’s (75)):

- (13) a. [To że tu jesteśmy] jest wszystkim wiadome.
 TO C here we are is to-everyone known
 ‘That we are here is known to everyone.’
 b. * [Że tu jesteśmy] jest wszystkim wiadome.

PT suggest that *że* might be an overt complementizer, not an instance of T moved to C, and *to* is a realization of T-to-C movement or an element that contains T-features or ϕ -features which are otherwise absent from the complementizer *że*.

In fact, the same data are also present in Russian, so I will briefly discuss them here. It is questionable whether *to* in (13a) can be viewed as part of the embedded CP; it is usually treated as a “pronominal correlative” of the embedded clause (see section 1.4.2 for a discussion of pronominal correlatives), that is, a DP (strictly speaking, a single D) occupying the subject position of the matrix clause, to which the embedded clause is adjoined (I give the second variant of the structure in the Russian example (14a)). In the latter case, it is possible to argue that in (13a) as well as in its Russian counterpart (14a) the embedded clause simply does not occupy the subject position, rather than that its complementizer is changed to allow it to occupy it. In any case, what (13b) and (14b) show is that neither Russian *čto* clauses nor Polish *że* clauses can occupy the subject position.

- (14) a. To, [čto my zdez’,] menja raduet.
 that that we here me pleases
 ‘It pleases me that we are here.’
 b. ?* [Čto my zdes’,] menja raduet.

To be precise, (14b) can be grammatical, but only with a marked stress pattern that shows that the embedded clause is topicalized and not in the subject position. Thus, it is possible to make a general claim that Russian and Polish complement clauses cannot appear in the subject position.³

However, the explanation of the contrast between English and Slavic, as well as of the English *that*-omission data, appears to be a weak spot in PT's argumentation. The analysis outlined above stops working once we discard the idea of feature deletion, which is the path taken in the most recent developments of the theory (Chomsky, 2005, 2006), because if features do not delete we cannot distinguish between the embedded CPs in (11c) and (11d) anymore. I will not appeal to feature deletion in my analysis, either; it seems a superfluous device if as a result of the Agree process unvalued features get valued and therefore do not "offend" the system any more. In (Pesetsky & Torrego, 2004b), PT propose a new account of the *that*-omission phenomenon based on a new variation of the feature deletion mechanism; however, as I will demonstrate below, this newer implementation also turns out to be flawed.

Pesetsky and Torrego 2004a

In (Pesetsky & Torrego, 2004a), PT elaborate on their theory of temporal features. The main contribution of this paper is the idea that not only nominative, but all instances of structural case – that is, both nominative and accusative in English – are actually instances of *uT* on D. Thus, not only subjects, but also objects bear an uninterpretable T feature and need to check it. PT propose that "*uT* on a complement of V enters an Agree relation with an instance of T that is structurally lower than the main tense of the sentence" (p. 496) – a head they call T_o ("object" T), in contrast to the main tense head T_s ("subject" T). This results in the following structure (PT's (16)):

$$(15) \text{ Subj } T_s [{}_{VP} v T_o [{}_{VP} V \text{ Obj}]]$$

The head T_o , just like T_s , bears uninterpretable ϕ -features and acts as a probe for a goal with a set of interpretable ϕ -features. However, as PT argue, it can only agree with a goal which also bears an instance of T (in this requirement, verbal T_o differs from instances of T found inside DPs and APs – heads which, in PT's view, serve to enable complementation for these categories). It follows that all verbal complements have to bear an instance of the T feature at the time when they are merged to V. Later, if the T feature in question is uninterpretable, it can and must delete, which is in line with the account of the *that*-omission asymmetry in (Pesetsky &

³However, as discussed further below, see (Koster, 1978) and (Stowell, 1981) for arguments in favour of the view that subject clauses are in fact always in a topicalized position and not in Spec,TP.

Torrego, 2001): the uT on the subject moved to CP will have deleted by the time when the subject clause is attracted to the matrix Spec,TP.

In this thesis, I will not adopt this modification of PT's theory and will not postulate the presence of a second T head. What troubles me is that the T₀ head seems to implement the function of object agreement, and basically does not do anything else (PT do make some suggestions about its possible semantic contribution, but do not say anything definite). I prefer to adopt the position of (Chomsky, 1995), who argues extensively against postulating a separate AgrO head in the structure, showing that the v head can perform the same functions and that AgrO can (therefore should) be dispensed with.

Pesetsky and Torrego 2004b

In their latest paper, PT propose a view on Agree that I will rely on in this thesis. They argue against the theory of Agree in (Chomsky, 2001, 2004), and in fact, their modifications are adopted in (Chomsky, 2005, 2006). The starting point is the view of the Agree operation and the inventory of features in the language expressed in (Chomsky, 2001, 2004). Linguistic features are distinguished in two ways: valued vs. unvalued and interpretable vs. uninterpretable. According to PT, these two properties of features are not identical. This is explained using the Latin sentences in (16) (PT's (1)) as an example:

- (16) a. Haec puella Romana
 this-Nom.Fem.Sg. girl-Nom.Fem.Sg. Roman-Nom.Fem.Sg.
 ambulat.
 walks-3.Sg.
- b. Hae puellae Romanae
 these-Nom.Fem.Pl. girls-Nom.Fem.Pl. Roman-Nom.Fem.Pl.
 ambulant.
 walk-3.Pl.

Consider the feature “feminine” on the noun *puella* and on the demonstrative *haec* and adjective *Romana* in (16a). PT assume that the noun comes out of the lexicon with the gender feature already specified as feminine; this means that this feature is valued. By contrast, the demonstrative and the adjective come out of the lexicon with their gender features lacking a specific value – i.e. unvalued, and get their “feminine” value in the course of the sentence derivation, due to agreement with the noun. On the other hand, the distinction between interpretable and uninterpretable features refers to the fact that some features make a semantic contribution to the interpretation on the item they belong to, while others do not. For example, the feature “number” on the noun *puellae* in (16b) will qualify as interpretable, because it serves to encode the fact that the sentence is

about multiple girls, while the same feature on the verb *ambulant* does not contribute anything to its interpretation.

So far, if we look at these and other features in the examples given, it seems that uninterpretable features are also unvalued, and vice versa. Indeed, this is what (Chomsky, 2001) assumes to hold universally, as is demonstrated by his Valuation/Interpretability Biconditional (PT's (3)): "A feature F is uninterpretable iff F is unvalued". This claim is motivated by the consideration that syntax cannot see whether a certain feature is interpretable or not because it cannot estimate the semantic contribution of an item; however, it can see whether or not a feature has been assigned a value. What triggers Agree is an unvalued feature (probe) which scans its c -command domain for a valued instance of the same feature (goal), and gets valued and interpreted if it finds one.

According to PT, another important property of Chomsky's (Chomsky, 2001) Agree mechanism is that once valuation has occurred, it is impossible for syntax to see that the valuation of a certain feature F_1 is due to agreement with another feature F_2 – "[t]here is no *link* established between F_1 and F_2 " (p. 3). In other words, the sole purpose of Agree is to get rid of the unvalued feature, and after this result is achieved, the process that led to it is no longer visible because it "does not interest" the system any more. PT term this approach the "assignment" version of Agree, because it is completed after the value of F_2 is assigned to F_1 .

(Pesetsky & Torrego, 2004b) depart from these assumptions to create their own model of Agree, which differs crucially from Chomsky's in that it sees Agree as a means to establish a link between the two features that take part in the agreement process – and, ultimately, between the two linguistic items that bear these features. They term their version of Agree "feature sharing". In contrast to the "feature assignment" approach to Agree, PT's system "remembers" completed agreement steps. The formulation of this version of Agree is given in (17) (PT's (5)):

(17) Agree (Feature sharing version)

- i. An unvalued feature F (a *probe*) on a head H at syntactic location α (F_α) scans its c -command domain for another instance of F (a *goal*) at location β (F_β) with which to agree.
- ii. Replace F_α with F_β so that the same feature is present in both locations.

PT use the term *occurrence of a feature* to describe identical features which have not (yet) undergone Agree, and the term *instance of a feature* to describe the manifestations of the same feature in distinct locations after the agreement process has taken place. An important consequence of the feature sharing approach is that a sequence of multiple instances of one and the same feature can be formed; for example, if after an unvalued feature FA has received its value from a valued feature FB, it itself values another

unvalued occurrence of the same feature FC, the feature F gets shared by three locations. Of course, the same sequence of feature instances would be present in the “assignment version” of Agree, but it would not have any importance because the three instances of F would not be connected to one another. This difference matters in case agreement happens between two unvalued features. As PT note, such a process is “either vacuous or impossible” under the assignment view – it is impossible if the goal of Agree must necessarily be valued, and if it is made possible it remains vacuous because it does not assign any value to any feature (for lack of value). However, under the feature sharing view an Agree operation between two unvalued features is not vacuous at all, because it replaces two distinct occurrences of a feature F with one occurrence of F shared between two different locations. If one of the instances of F then agrees with a valued occurrence of the same feature, both of the agreeing unvalued feature instances will get a value. It follows that only one valued instance of a feature will suffice to give values to all other instances of this feature linked up by Agree. It is clear that Agree seen this way is a very powerful syntactic device, and it is crucial for Reuland’s (Reuland, 2001, 2005) theory of binding as well as for my SOT theory.

The second modification of (Chomsky, 2001) that PT include in their theory of Agree is challenging the “Valuation/Interpretability Biconditional”. They argue that it is in fact puzzling that such distinct properties of linguistic items as valuation and interpretability should necessarily be paired. Instead, they propose two independent classifications of features based on the two properties, which yields four combinations. Two of the resulting feature types – uninterpretable unvalued features and interpretable valued features – are the only ones possible under the view in (Chomsky, 2001). To these, PT add two new types: uninterpretable valued features and interpretable unvalued features. As for Chomsky’s argument in favour of the Biconditional, they show that it is not inconsistent with their proposal: the fact that syntax cannot recognize uninterpretable features as such (that is, as potentially offending elements) means that only unvalued features can be probes for Agree; however, it does not follow that they should necessarily be uninterpretable as well.

From the discussion so far, it seems that if syntax cannot distinguish between interpretable and uninterpretable features, then it simply does not care about them. Indeed, the feature sharing approach to Agree formulated above relies exclusively on the notion of valuation and does not say anything about interpretability. However, PT do not discard the view that uninterpretable features must be interpreted. They rely on the thesis of Radical Interpretability by (Brody, 1997), which they state as follows (their (15)): “Each feature must receive a semantic interpretation in some syntactic location”. Thus, no uninterpretable feature is allowed to survive until the transfer of the derivation to the semantic (C-I) interface. Consequently, uninterpretable feature occurrences have to enter an Agree

relation with interpretable ones; a feature without a single interpretable instance will not receive any semantic interpretation and so it will cause the derivation to crash at the interface. We now have a final modification of the feature sharing hypothesis: Agree links up two or more instances of a feature, among which there has to be at least one valued and at least one interpretable instance; this will suffice to give valuation and interpretation to all other agreeing feature instances.

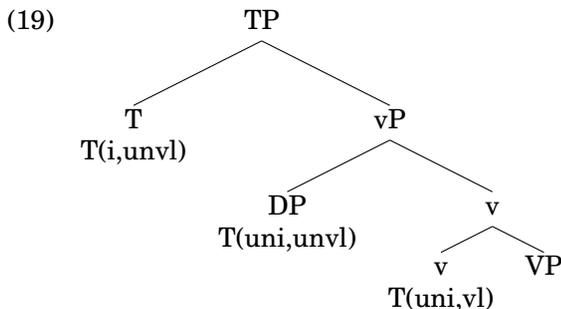
As a result, in PT's system, there are two types of features that can function as probes: uninterpretable unvalued features and interpretable unvalued features. The second type of probes, which is a new one for the theory, is exemplified by the T(ense) feature on the T head (termed Tns by (Pesetsky & Torrego, 2004b) to avoid confusion), while the same feature on the finite verb is taken to be, in contrast, uninterpretable but valued. This is how PT motivate this view: "If Chomsky (1957), Emonds (1976, 1978), Pollock (1989) and others are correct in positing a distinct Tns node as the locus of semantic tense interpretation, the theory must take cognizance of the fact that in many languages, the finite verb – not Tns itself – bears the morphology that makes tense distinctions. This means that T on the finite verb in such languages is an uninterpretable feature that participates in an Agree relation with T on Tns. Since Tns c-commands the finite verb, its T must be the probe in this relation. Consequently, T on Tns must be an interpretable feature that is unvalued and acts as a probe. Likewise, T on the finite verb must be an uninterpretable feature that is valued and acts as a goal." (p. 6) In note (9), PT note that they take the finite verb to be *v* rather than *V*, and refer to a fuller presentation of the material in their work in preparation, where they argue that the verbal domain is built in the following way: *V* comes from the lexicon from an uninterpretable valued T feature, and *v* with an uninterpretable unvalued T feature. Then the T feature on *v* acts as a probe that agrees with T on *V* and gets its value from it. Thus, three heads – T, *v* and *V* – end up bearing instances of the T feature.

Based on this approach to the T-feature, PT outline a proposal to account for raising constructions like the one in the English sentence (18) (adapted from their (16)). I will discuss it in some detail because I will use a similar mechanism in my own treatment of the SOT phenomena.

(18) Mary seemed to like the play.

In contrast to the analysis in (Chomsky, 1995, 2001) and to their own earlier analysis, which proposed that the ϕ -features of the T head act as a probe in the raising process, PT now claim that raising is driven by the T feature on the T head that enters into an Agree relation with the uT on the subject DP. In a "normal" finite clause, valuation of uT on the subject proceeds in the following steps: first, the interpretable but unvalued T feature on the T head agrees with the uninterpretable unvalued T feature on the subject DP. However, this agreement process does not result in valuation.

Therefore, the unvalued T feature on the T head probes again and agrees with the uninterpretable but valued T feature on *v*. Now both of the unvalued T features (on T and on DP) get a value, while the uninterpretable T feature on *v* gets interpreted, as shown in (19), simplified from PT's (17) (*i* stands for interpretable, *vl* for valued, *uni* for uninterpretable, *unvl* for unvalued):



PT propose that in a raising construction like (18), the lower *v* enters the derivation not only uninterpretable but also unvalued. This means that after the Agree operations shown in (19), the T feature on the T, subject DP and *v* heads of the lower clause is still not valued. It then receives its value from the matrix clause: first, the subject DP moves into the higher clause; then it is probed by the unvalued matrix T; finally, the matrix T probes for the valued instance of T on the matrix finite verb *v*, and all the instances of T in the sentence get valued. The semantic consequence of this process is that the infinitive is dependent on the higher tense for its interpretation. This last point – the idea of deriving the semantic dependency between two tenses in a sentence by postulating an agreement process between the two corresponding T(ense) heads – seems very insightful, and I will show how it can be used to encode temporal dependencies in a SOT configuration.

As was mentioned above, in their latest paper PT also propose a new account for the *that*-omission asymmetry in English. Instead of positing a timing restriction on the deletion of features, they modify the mechanism of feature deletion proposed in (Chomsky, 2001): they argue that only uninterpretable instances of a feature have to delete at the interfaces after they have been interpreted, while interpretable instances – no matter whether they start out as valued or unvalued – stay undeleted. In a sentence like (11c), repeated here as (20a), the feature T on C is interpretable – in fact, it is the head T itself in the shape of *that* that has moved to C, together with its interpretable T feature. However, in (11d), repeated in (20b), the feature T on the subject gets deleted as uninterpretable.

- (20) a. [That Sue will buy the book] was expected by everyone.
 b. * [Sue will buy a book] was expected by everyone.

So, the difference between the two embedded clauses remains the same: one CP contains an instance of the T feature by the time it is merged at the matrix Spec,TP, the other does not contain it. A further modification of the explanation is that it is the unvalued T feature on the matrix T that probes for the embedded C and attracts it into the subject position; naturally, the CP bearing an instance of T feature can be moved into this position, while a “T-less” CP cannot. This mechanism is especially attractive since it allows PT to dispense with the Match Condition on the probe-goal relations.

This analysis is quite elegant, however it is not without problems. First, postulating a deletion mechanism for uninterpretable features does not appear to be independently justified. In PT’s framework, unvalued features get a value through agreement and, once valued, do not make the derivation crash, just like the instances of the same feature which were valued to begin with. This is a simple one-step process; however, for getting rid of uninterpretable features, a two-step syntactic process is proposed. The first step is to establish an Agree relation with an interpretable instance of the feature. The second step is to delete the uninterpretable feature instances once the derivation reaches the semantic interface. However, the first step seems sufficient for preventing the potential harm that uninterpretable features can do to a derivation. According to the thesis of Radical Interpretability, a feature must receive semantic interpretation in *some* syntactic location – that is, not necessarily in *all* locations. Thus, we can suggest that the interpretable instance(s) of the feature can provide interpretation for all agreeing uninterpretable feature instances, just as a valued feature provides valuation; hence no “uninterpretable features” have to delete.⁴

More importantly, the analysis has a drawback that PT themselves notice. If the higher T feature enters into an agreement relation with the lower T feature (via the instance of T on the subject of the matrix clause), this means that all instances of T in the resulting structure – both in the matrix and in the embedded clause – must share a value. PT admit that such a configuration should always result in a semantic dependency between the tenses like the one seen in the raising construction in (18). To cope with this problem, PT propose that values of T are not specific tenses, like present, past or future, but simply “plus” and “minus” – or rather, a valued T is “plus”, while an unvalued is “minus”. Thus, sharing a value should not have any consequences for the temporal interpretation of different clauses. As for tenses, they “do not correspond to values of a grammatical feature, but constitute different sorts of *encyclopedic* information that may be associated with a T feature that has a positive value (i.e. [iT

⁴If these considerations are on the right track, the distinction between interpretable and uninterpretable features does not play a crucial role in the analysis. With this modification, PT’s approach to Agree becomes in principle compatible with theories that do not allow for existence of uninterpretable features at all. I thank Øystein Nilsen (p.c.) for pointing out to me the importance of this issue.

+])” (p. 17). PT illustrate this point by referring to lexical items like *dog* and *giraffe*, which share a positive value for the animacy feature but differ in their “encyclopedic specifications” (ESs). Applying this analogy to the verbal domain, PT assume that the temporal meanings like past or present are in fact ESs that are contributed by the temporal morphology on the finite verb. They are different for each clause containing a finite verb and cannot be influenced by agreement.

There are several reasons why this solution is not optimal. First, PT do not specify what exactly is the meaning of the “plus” and “minus” values for tense. In note (31), PT argue that this distinction does not coincide with the finiteness/non-finiteness dichotomy, because some infinitive clauses can be subjects and so have to have a “plus” value for tense. They suggest that prepositions might be elements characterized as “minus T”. This would mean that the +/-T opposition does not play any role in the the interpretation of verbal forms (since all of them are +T).

Besides, the term “ES” can refer to nothing other than the lexical meaning of a word, and indeed this is what *dog* and *giraffe* differ in; in other words, temporal ESs will have to be the “lexical meanings” of the temporal affixes. However, since PT argue that the locus of temporal interpretation is in T and not in v, the ESs contributed by v should “reach” T somehow, and the only way to implement it is by the T-v agreement. This makes agreement between T and v a phenomenon very different from the other instances of agreement found in the language, since lexical meanings are never transmitted by agreement. But in case of PT’s temporal agreement, along with the “plus tense” value, v will have to give to T its ESs (e.g., past or present). PT recognize this problem and claim that “although the ES of a feature is not *shared* by the feature’s various locations, it may be accessed at any of these locations”; however, the difference seems to be of a terminological character.

Furthermore, to explain the temporal interpretation of infinitival constructions like (18), PT have to assume that in some cases ESs *can* in fact be “transmitted” between two different clauses. In order to allow the embedded T to receive its interpretation from the matrix T, PT have to introduce an independent “Economy condition” on the accessibility of ESs (without it, deriving any interpretation for the embedded T becomes tricky; see (Pesetsky & Torrego, 2004b) for details).

My final remark concerns my own proposal which builds on PT’s approach to agreement relations. In the following chapters of my dissertation, I will show that an Agree relation between two finite T heads leading to a temporal dependency does in fact occur in the language, namely in case of sequence of tense. This brings me to the conclusion that the mechanism of temporal agreement is indispensable to the theory, and cannot be done away with.

If we stick to this view, it is definitely not desirable to maintain an analysis where the T feature on the matrix T head is checked by the T

feature on the subject clause C, since that would always lead to a temporal dependency. However, this problem is solved if we adopt the approach defended in (Koster, 1978; Stowell, 1981), who argue (as already mentioned in footnote 3 on page 27) that subject clauses in the proper sense do not exist, and the clauses that appear to be in subject position are in fact in a topicalized position. The subject position could in this case be occupied by an expletive, which is obligatorily dropped when the embedded clause is fronted. However, in that case, the contrast between (20a) and (20b), as well as the crosslinguistic contrast illustrated by (13)-(14) on pages 26-27, remain unexplained; this issue has to be put aside for future research.

The last thing I will mention in this section are PT's (Pesetsky & Torrego, 2004b) views on wh-movement, also derived from their theory of Agree as feature sharing. They argue that wh-movement is driven by different occurrences of the same feature on the wh-word and on C – the feature they call Q. On the C head, the Q feature is interpretable but unvalued, and it probes for the Q feature on the wh-word, which is uninterpretable but valued (attracting the corresponding head in languages with wh-movement). As in the case of T-v agreement, the unvalued feature occurrence receives its value via the agreement process, and the locus of interpretation is the higher feature instance (in C). Further, PT argue that an uninterpretable and unvalued instance of the Q feature can be found on “intermediate” C heads supporting successive cyclic movement. Long-distance wh-movement then proceeds in the following way: first, the uninterpretable and unvalued instance of Q on the embedded C probes for the interpretable unvalued Q on the wh-word and attracts it; however, the Q feature is still not valued, and the wh-word gets attracted further until it reaches the matrix C with valued Q. Then all the lower instances of Q get valued.

2.1.4 Summary

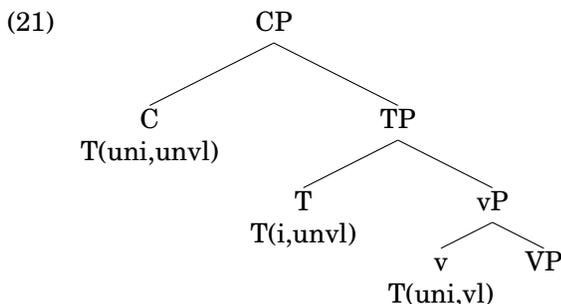
This section dealt with the notion of the C-T system. After summarizing contemporary views on phrase structure, I outlined the main claims of Phase Theory, in particular the important claim that CP is a phase and TP is dependent on it. Further, I analysed the model of the functional layer of the clause developed by (Pesetsky & Torrego, 2001, et seq.). I will adopt the following ideas from PT's theory:

1. The C head in English bears an uninterpretable unvalued T feature. This seems to me a logical way to implement the general idea of the connection between C and T. It will not be important for my work how exactly this feature is satisfied in each particular construction; what is important is that an Agree relation between C and T is eventually established.
2. Grammatical features are divided into interpretable vs. uninter-

pretable, on the one hand, and valued vs. unvalued, on the other hand. Agree is always triggered by unvalued features.

3. Agree is feature sharing; syntax “remembers” steps between agreeing instances of the same feature. Therefore, several unvalued feature instances linked by an Agree relation can all get valued if one of them agrees with a valued occurrence of the same feature. At least one valued and one interpretable instance of a feature is needed for all agreeing instances to count as valued and interpretable.

Finally, it is important for me to trace the valuation of the T feature in a clause. I assume that instances of this feature are found on C, T, v (and presumably also V; however, it is not going to be crucial for my purposes so I omit V here). In my derivations I will ignore the T feature on the subject DP, since it is not relevant for my analysis. I present the resulting clause structure in (21):



2.2 Binding theory

This section is dedicated to Binding Theory. Since I propose to use a binding mechanism for temporal dependencies, it is important for me to define what version of the binding theory I rely on and what governs my choice. First, in section 2.2.1 I will outline the main notions of the classical Binding Theory based on (Chomsky, 1981). In section 2.2.3, I will present a summary of a more recent take on the BT by (Reinhart, 1983, et seq.), (Reinhart & Reuland, 1993). Finally, in section 2.2.4 I will present the version of the BT by (Reuland, 2001, 2005). The binding mechanism proposed in this work will be the most important for me, because I will adopt the principle of Binding via building of agreement chains.

2.2.1 Classical binding theory

Binding theory (hereinafter BT) deals with conditions for sentence-internal anaphora; it studies syntactic constraints on the interpretation of nominal expressions, in particular, anaphoric elements – expressions that depend for their interpretation on other NPs (their *antecedents*). The first

version of the Binding theory – I will call it the classical Binding Theory – was developed in (Chomsky, 1981) (with an earlier version presented in (Chomsky, 1980)) as part of the Government and Binding Theory. It relies on the assumption that all NPs are generated with freely assigned indices, which are responsible for representing dependencies between nominal expressions; these indices may be identical on different NPs, which results in coindexation, so that the same value can be assigned to multiple NPs. Sentence-internal coindexation is constrained by syntactic principles which are different for different types of NPs. Classical BT divides NPs into three types: anaphors, pronouns and referring expressions (R-expressions). The first two classes are anaphoric elements. The anaphor class includes reflexives like *herself* and reciprocals like *each other*, and the pronoun class includes personal pronouns like *she*. The rest of NPs are classified as referring expressions, that is, they do not depend on an antecedent for their interpretation.⁵ The behaviour of the three types of elements is exemplified by sentences (22a)–(22i).

- (22) a. Mary_i likes herself_i.
 b. *I_i like herself_i.
 c. Mary_i's sister_j likes herself_{*i/j}.
 d. *Mary_i says that John likes herself_i.
 e. Mary_i likes her_{*i/j}.
 f. Mary_i says that John likes her_i.
 g. She_{*i/j} likes Mary_i.
 h. She_{*i/j} thinks that John likes Mary_i.
 i. When she_i was at school, Mary_i liked John.

Sentences (22a)–(22b) show that the element *herself* can only be coindexed with its antecedent if both expressions have matching features – namely, person, number and gender. However, this is not a sufficient condition for linking *herself* to another NP: as the contrast between (22a) and (22c) shows, the structural relation of c-command between the antecedent and the anaphor is required. If these conditions are satisfied, the relation of binding is established between the anaphor and its antecedent. The definition of binding is given in (23), while the relevant formulation of c-command is given in (24):⁶

- (23) α binds β iff α and β are coindexed and α c-commands β (Reinhart, 2000).

⁵In BT (starting with (Chomsky, 1981)), the term *referent* is normally used to denote not necessarily an object in the real world, but rather “a fixed value in some domain” ((Grodzinsky & Reinhart, 1993, p. 72)); this use was adopted in much later work, and I will stick to it as well. Still, obviously this use of the term ‘referring’ is an overgeneralization, since it abstracts away from non-referential NPs.

⁶It should be noted that classical BT deals with *A-binding* only, that is, binding by an antecedent in an A-position; here I use the term *binding* to mean *A-binding*.

- (24) α c-commands β iff the branching node most immediately dominating α also dominates β (Reinhart, 1983, p. 18).

Thus, in (22a) *herself* is bound by *Mary*; (22b) demonstrates that *herself* has to be bound, otherwise the sentence becomes ungrammatical. However, (22d) shows that if the binder of an anaphor (here, the NP *Mary* in the matrix clause) is not “close enough”, this also results in ungrammaticality. In other words, an anaphor must be bound in a certain local domain, termed a ‘governing category’ in (Chomsky, 1981). (According to (Chomsky, 1981), this domain was supposed to contain a governor for the anaphor; in addition, it had to include an element termed “accessible SUBJECT” which could be either a subject NP or agreement (AGR)).

Sentences (22e) and (22f) show that pronouns exhibit the opposite behaviour in comparison to anaphors: they must not be bound in their local domain, however they can be bound by a suitable antecedent outside it. In contrast to anaphors, pronouns do not necessarily have to be bound: the impossibility of binding *her* in (22e) does not render the sentence ungrammatical, because it can be saved if *her* is left unbound, or *free*. Thus, the only condition on the occurrence of pronouns is that they must not be bound locally. Finally, (22g) and (22h) demonstrate that a R-expression like *Mary* must always be free; to be sure, it can corefer with another nominal expression, but only a non-c-commanding one, as can be seen from (22i). The rules outlined above were summarized in (Chomsky, 1981, p. 188) as the binding principles, or conditions:

- (25) (Principle A) An anaphor is bound in its governing category.
 (Principle B) A pronominal is free in its governing category.
 (Principle C) An R-expression must be free.

2.2.2 Syntactic binding and logical syntax binding

It is important to distinguish syntactic binding, discussed above, from the logical syntax notion of binding. In the latter sense, binding means binding of a variable by an operator. This operation can be illustrated by the bound reading of a sentence like (26a), represented in (26b):

- (26) a. Every girl thinks that she is smart.
 b. Every girl (λx (x thinks that x is smart)).

The pronoun in (26a) is interpreted as a variable, which can be bound by an operator or left free. In (26b), the variable is bound by the λ -operator.

In contrast, in BT (e.g., Chomsky, 1981), one NP is supposed to bind another (thus, in (26a), *every girl* binds *she*), which does not make sense from the logical binding point of view, since the binder is not an operator. As (Reinhart, 2000) puts it, “[l]ogical binding is a relation between operators and variables, and not between arguments, but syntactic binding is

a relation between variables (indices), i.e. between arguments” (Reinhart, 2000, p. 2). In order to eliminate this contradiction, Reinhart (2000) offers a logical syntax-based definition of argument binding (A-binding):

- (27) α A-binds β iff α is the sister of a λ -predicate whose operator binds β . (Reinhart, 2000, (11))

By this definition, *every girl* can bind *she* in (26a), since it is a sister of the λ -predicate whose operator binds the pronoun.

2.2.3 A new version of BT: (Reinhart, 1983, 1986; Reinhart & Reuland, 1993)

The classical BT as formulated in the three binding principles had considerable empirical coverage, however it faced several serious problems. One of them was the exact definition of the local domain in which anaphors were always bound and pronouns were free. This turned out very difficult to capture. The definition of governing category in (Chomsky, 1981) later underwent several modifications, but exceptions that did not fit it always remained. Another major problem was that the binding principles required a complementary distribution of anaphors and pronouns, but this requirement turned out to be too strict: for example, long-distance binding of anaphors appears to be widespread. All this led to revisions of the classical BT. In what follows, I will concentrate on the implementation of the binding theory offered in (Reinhart, 1983, 1986) (summarized in (Grodzinsky & Reinhart, 1993)), and in (Reinhart & Reuland, 1993).

Reinhart (1983, 1986); Grodzinsky & Reinhart(1993)

(Grodzinsky & Reinhart, 1993) (hereinafter GR) summarize the modifications to the BT developed in the preceding work by (Reinhart, 1983, 1986); I will follow their summary to present Reinhart’s ideas. Reinhart’s theory is based on the distinction between two types of anaphoric relations present in the language: binding, that is, variable binding in the logical sense, and coreference, a process that “defines a relation between linguistic structures and their potential use for the purpose of expressing referential intentions” (Grodzinsky & Reinhart, 1993, p. 72). For example, pronouns in (28) and (29) (GR’s (4a) and (5a)) are linked to their antecedents by coreference, and in (30) (GR’s (6a)) binding takes place. However, not all syntactic configurations allow for binding, as (31) (GR’s (7a)) demonstrates:

- (28) Lucie_i adores her_i friends.
 (29) Most of her_i friends adore Lucie_i.
 (30) Every actress_i adores her friends_i.
 (31) * Most of her_i friends adore every actress_i.

The classical BT puts both binding and coreference in the grammar; in contrast, Reinhart claims that grammatical BT should deal with binding in the first place, leaving coreference to the interpretative system. Thus, anaphors are elements that always have to be bound.⁷ As for pronouns, Reinhart argues that they can enter binding relations (in the sense of variable binding) not only with a quantified expression, but also with any other type of antecedent. This means that whenever a pronoun is coindexed with a *c*-commanding antecedent, two variants of interpretation are in principle possible: one produced by binding and another produced by coreference, as illustrated in (32) (GR's (9)):

- (32) Alfred_i thinks he_i is a great cook.
 a. Alfred (λx (x thinks x is a great cook))
 b. Alfred_i (λx (x thinks he_i is a great cook))

The two operations yield an identical result in the general case, however they can be disambiguated. In a sentence like (33) (GR's (10)), the strict reading (meaning that Felix considers Alfred a great cook) results from coreference in the first two clauses, and the sloppy reading (Felix considers himself a great cook) results from binding. A similar ambiguity is obtained in (34) (GR's (11)) – it can imply either that other people do not consider Alfred a great cook, or that they do not think of themselves as great cooks.

- (33) Alfred_i thinks he_i is a great cook, and Felix does too [e].
 (34) Only Alfred_i thinks that he_i is a great cook.

To sum up, binding is available to both anaphors and pronouns, but the coreference option is open only for the latter. The precise conditions for binding of anaphors and pronouns is what BT should provide. Obviously, coreference is not completely arbitrary, either, but Reinhart argues that it is not governed by syntax; rather, it is established at the discourse level. To rule out Condition B violations, Reinhart argues that coreference is excluded in cases where one of the coreferring elements can be replaced by a bound variable yielding an identical interpretation; this is captured by the intrasentential coreference rule (GR's (20)):

- (35) *Rule I: Intrasentential Coreference*
 NP A cannot corefer with NP B if replacing A with C, C a variable A-bound by B, yields an indistinguishable interpretation.

Rule I in (35) rules out (36) (GR's (18a)) because *her* could be replaced by a variable bound by the NP *Lucie* with indistinguishable interpretation. It follows that in sentences like (32), the coreference interpretation of the

⁷This view is modified in (Reinhart & Reuland, 1993), where it is argued that the coreference option is also available to anaphors.

pronoun is only available when it can be distinguished from the bound interpretation (for instance, in (33)). Interestingly, coreference is possible in (37) (GR's (19d)), since it produces an interpretation which is different from the binding interpretation and is more adequate in the sentence: what Ann and Bill have in common is the property of admiring Bill, not the property of self-admiration.

(36) Lucie_i adores her_k.

(37) I know what Ann and Bill have in common. She thinks that Bill is terrific and *he* thinks that *Bill* is terrific.

It should be noted that in (Reinhart, 2000), the distinction between binding and coreference is modified to include cases where no referential antecedent is available but an ambiguity between a bound and a "not-bound" reading still arises, as in (38) ((Reinhart, 2000)'s (7), taken from (Heim, 1998)):

- (38) a. Every wife thinks that only she respects her husband.
 b. Every wife (λx (x thinks that [only x respects x's husband]))
 c. Every wife_i thinks that only she_i respects her_i husband.

To cover readings like (38c) as well as (32b), Reinhart uses the term *covaluation* instead of *coreference*, and gives it the following definition (her (12)):

(39) Covaluation:

α and β are covalued iff neither A-binds the other and they are assigned the same value.

The exact details of the analysis in (Reinhart, 2000) are not crucial for me, and I will not go into them here; however, I will adopt the terminology introduced in this work, and will use the term *covaluation* to refer to discourse-level anaphora resolution.

Reinhart & Reuland (1993)

In the work mentioned above (Reinhart, 1983, 1986; Grodzinsky & Reinhart, 1993), the binding principles A and B were left unmodified. (Reinhart & Reuland, 1993) (hereinafter RR) propose new binding conditions which fare better with the empirical challenges mentioned in the beginning of this section. They start by redefining the typology of anaphoric expressions. The standard BT divides all anaphoric expressions into pronouns and anaphors. RR propose a further division of anaphors into two groups: simplex expression anaphors (SE anaphors), which can normally function as long-distance anaphors (such as the Dutch *zich*, the Norwegian *seg*, etc), and complex expression anaphors, termed SELF anaphors by RR, which

are usually interpreted locally (such as the Dutch *zichzelf*, English *herself*, etc). Thus, we arrive at three classes of anaphoric expressions instead of two. The classification is based on two properties. The first one is the property of referential independence (+/-R), that is, the ability to project an independently interpreted argument; this property distinguishes pronouns, which can be referentially independent, from anaphors, which are referentially defective and have to be bound. The second property is the ability of the anaphoric element to function as a reflexivizer; reflexivization is defined as “imposing identity on two arguments of a predicate” (p. 659), which is a function available only to SELF anaphors. The resulting classification is given in (40) (RR’s (4)):

	SELF	SE	Pronoun
(40) Reflexivizing function	+	–	–
R(eferential independence)	–	–	+

RR argue that only the reflexivizing function of anaphoric elements is governed by the BT, while conditions of their use that depend on their referential properties fall under chain theory. Furthermore, they argue that the binding conditions are in fact conditions on the behaviour of reflexive predicates, rather than on the use of pronominals. They propose that reflexivity must always be licensed in the language, and that it is done by marking either the head of the reflexive predicate (for example, a verb) or one of its arguments. Intrinsically reflexive predicates are marked as reflexive in the lexicon, and they can only be used reflexively, though the absorbed theta-role has to be realized overtly in some languages. Extrinsically reflexive predicates are transitive predicates that are made reflexive by reflexive-marking one of their arguments (with a SELF anaphor). The definitions are summarized in (41) (RR’s (11’)):⁸

(41) *Definitions*

- a. A predicate is *reflexive* iff two of its arguments are coindexed.
- b. A predicate (formed of P) is *reflexive-marked* iff either P is lexically reflexive or one of P’s arguments is a SELF anaphor.

RR propose to restate the Condition B of the BT in terms of the obligatory marking of reflexivity (RR’s (12’)):

(42) *Condition B*

A reflexive predicate is reflexive-marked.

In other words, if a predicate is reflexive, reflexivity has to be marked somehow: either in the lexicon (on the predicate head itself), or in the syntax (on one of the arguments).

⁸The first version of the definitions (RR’s (11)) is somewhat more elaborate, then it is simplified for the sake of exposition; I’m giving the simplified version here.

The new Condition B efficiently rules out pronouns in positions where they are (speaking in terms of the classical Condition B) bound in their governing category, like in (43a) and (43c) (RR's (13a) and (13c)): in both of these sentences, the two arguments of the predicate are coindexed, so the predicate is reflexive, but it is not intrinsically reflexive, so the reflexive marking has to appear on its argument. However, the pronoun *he/him* is not allowed to act as a reflexive marker, which leads to the violation of Condition B as stated in (42). By contrast, in the grammatical (43b) (RR's (13b)) the predicate is reflexive-marked by the SELF anaphor *himself*.

- (43) a. *Max₁ criticized him₁.
 b. Max₁ criticized himself₁.
 c. *Max₁/he₁ criticized Max₁.

The occurrence of SE anaphors in the same position is also ruled out, as shown in (44a) (RR's (17a)). In some cases, though, SE anaphors are allowed to be bound by a coargument of the verb ((44b), (44c), RR's (19a), (19b)). The explanation is the following: "A verb like *schamen* 'shame' is intrinsically reflexive, as witnessed by the fact that it cannot take any object distinct in reference from the subject. Verbs like *wassen* 'wash', which do allow a distinct object, are listed twice in the lexicon, both as reflexive and as a nonreflexive; their reflexive entry allows the SE anaphor, and their transitive entry occurs with a SELF anaphor" (p. 665). So, an argument of an intrinsically reflexive predicate has to be realized as the SE anaphor *zich* in Dutch; *zich* is allowed to occur with an intrinsically reflexive predicate because the predicate is already reflexively marked in the lexicon, but it cannot occur with an extrinsically reflexive predicate like *haten* 'hate' since it cannot reflexive-mark an argument.

- (44) a. *Max haat zich.
 Max hates SE
 b. Max wast zich.
 Max washes SE
 c. Max schaamt zich.
 Max shames SE
 'Max is ashamed.'
 d. Max wast zichzelf.
 Max washes himself

As for the Condition A of the BT, which deals with the distribution of anaphors, RR propose that it should also be restated in reflexivity terms. The new Condition A is given in (45) (RR's (12')). It should be noted that the two Conditions cannot be collapsed into one "two-way" condition linking reflexivity to reflexive-marking, because RR argue that different types of predicates are involved in the two conditions: Condition A is about syntactic predicates, while Condition B is about semantic predicates.

(45) *Condition A*

A reflexive-marked predicate is reflexive.

This condition requires that a predicate whose argument is reflexive-marked (i.e., is a SELF anaphor) be reflexive. Hence, it correctly predicts the contrasts between (43b) (repeated here as (46a)), (46b) (RR's (22b)) and (46c) (RR's (22a)):

- (46) a. Max₁ criticized himself₁.
 b. * Five tourists talked to myself in the room.
 c. There were five tourists in the room apart from myself.

In the grammatical (46a), the predicate is reflexive-marked, and its arguments are coindexed, so it is reflexive. In (46b), on the other hand, the predicate *talked* cannot be reflexive (because of the feature mismatch between its arguments), so it is not licit to reflexive-mark its argument with a SELF anaphor. As for (46c), which is problematic for the classical Condition A, it does not involve a reflexive-marked predicate: although it does contain a SELF anaphor, the anaphor is not an argument of the verb. RR's Condition A only specifies the conditions on reflexive-marking, and does not say anything about other occurrences of SELF anaphors, so it does not rule out (46c).

The conditions A and B define the difference in distribution of SELF anaphors, on the one hand, and SE anaphors and pronouns, on the other hand. Further, RR argue that the distribution of SE anaphors as compared to pronouns is not governed uniquely by the BT; rather, it falls partly under the theory of A-Chain formation (Chain Theory). An A-Chain is defined broadly as “any sequence of coindexation that is headed by an A-position and satisfies antecedent government (...)” (p. 693); thus, an anaphoric expression together with its c-commanding antecedent represents an instance of an A-chain. RR propose the following syntactic conditions on well-formed A-chains (their (80)):

(47) General condition on A-chains

A maximal A-chain ($\alpha_1 \alpha_n$) contains exactly one link – α_1 – that is both +R and Case-marked.

The +R property, which serves to distinguish anaphors from pronouns, is defined in (48) (RR's (81)):

- (48) An NP is +R iff it carries a full specification for ϕ -features and structural Case.

Thus, only a -R expression can be the foot of an A-chain. Being -R, SE anaphors can implement this role, while pronouns cannot, which accounts for the contrast in (49) (RR's (89a,b)):

- (49) a. Willem₁ schaamt zich₁.
 Willem shames SE
- b. *Willem₁ schaamt hem₁.
 Willem shames him

2.2.4 Binding and agreement (Reuland, 2001, 2005)

(Reuland, 2001, 2005) further develops the line of research taken in (Grodzinsky & Reinhart, 1993; Reinhart & Reuland, 1993), addressing again the question of the distribution of pronouns vs. SE anaphors. As shown in the previous section, in (Reinhart & Reuland, 1993) this problem was solved by means of the condition on A-chains given in (47) above. However, this solution was based on coindexation chains – a notion eliminated from the theory in (Chomsky, 1995, et seq.). The Inclusiveness Condition, introduced in the Minimalist program, ensures that “any structure formed by the computation is constituted of elements already present in the lexical items selected” (Reuland, 2001, p. 440). Therefore, a theory of binding can no longer rely on coindexation, since no new objects such as indices can be added to the lexical items in the course of the derivation; so, a new mechanism governing the relation between the anaphor and its antecedent in examples like (49a) has to be found.

Recall that GR propose a ranking between binding and coreference formulated in the intransentential coreference Rule I (given in (35) on page 40). This rule can be understood (with empirical reasons, as has been shown in subsequent work, see e.g. (Chien & Wexler, 1990)) as an economy condition: binding is “less costly” than coreference, so, other things being equal, it is preferred. (Reuland, 2001) applies the same reasoning to the distribution of pronouns vs. SE anaphors: a sentence like (49a) must be preferred over (49b) because for some reason, the relation between the SE anaphor and its antecedent in (49a) is “less costly” than the binding relation in (49b). This intuition is formulated in (50) ((Reuland, 2001)’s (25)), modelled after the Rule I in (35):

- (50) NP A cannot be A-bound by NP B if replacing A with C, C an NP such that B R C, yields an indistinguishable interface representation.

The task is now to determine what kind of relation R is. (Reuland, 2001) follows RR in proposing that it is a chain relation. Under minimalist assumptions, the only way to establish dependencies between syntactic objects is by the operations of movement and feature checking/deletion, so only these mechanisms can be used in chain formation. (Reuland, 2001) introduces the notion of Chains – objects formed by feature checking and deletion – in addition to (generalized) chains, as defined in RR. The definition of Chain is given in (51) ((Reuland, 2001)’s (41)):

(51) *Chain*

(α , β) form a Chain if (a) β 's features have been (deleted by and) recovered from α , and (b) (α , β) meets standard conditions on chains such as uniformity, c-command, and locality.

This definition by itself cannot be made to work for antecedent-anaphor dependencies since in (Chomsky, 1995) it is assumed that a checking/deletion dependency can be established only if one occurrence of the feature(s) is uninterpretable, and the features of both NPs in a sentence like (49a) are interpretable. This follows from the principle of recoverability of deletion (PRD): a feature can be deleted only if no information is lost in the process; whatever is deleted has to be recoverable. However, (Reuland, 2001) argues that interpretable features can be deleted if this does not violate the PRD, that is, if they are fully recoverable from the interpretively equivalent features that delete them, which is indeed the case for the ϕ -features on a SE anaphor – they can be recovered from their antecedent. In contrast, the features on the pronoun in (49b) cannot be deleted because in contrast to SE anaphors, pronouns are specified for number; (Reuland, 2001) claims that “[n]umber is like a lexical element; different occurrences of it in the numeration make independent contributions to interpretation” (p. 458). Since in his system, the checking operation involves not separate features but feature bundles (e.g. category, number, person and gender on an NP), the impossibility to delete number makes Chain formation illicit.

This approach is further elaborated in (Reuland, 2005), where the relation between the anaphor and its antecedent is explained not in terms of Chain formation but in terms of agreement. In my own proposal, I will use the resulting model to implement the dependency between temporal morphemes, so I will review it in some detail here and will refer to this presentation in the following chapters.

(Reuland, 2005) adopts the following claims made in (Pesetsky & Torrego, 2004b): features are divided into interpretable/uninterpretable, on the one hand, and valued/unvalued, on the other hand; Agree is feature sharing, and a feature chain is licit if it has at least one valued and one interpretable instance; nominative case is an instance of unvalued, uninterpretable T(ense) on DP. He proposes that SE anaphors like the Dutch *zich* carry unvalued interpretable number and gender features. This means that in the course of the derivation, they have to be valued somehow, and they get the required values from their antecedent. In the usual case, that will be the subject (or, in Reuland's terms, EA, external argument) of the clause, as in the Dutch (52) (Reuland's (5)):

- (52) *Iedere professional voelde [zich aan de kant geschoven].*
 Every professional felt himself to the side pushed

A simplified structure of (52) is given in (53) (Reuland's (6)):

- (53) [Tns [EA [v* [V SE ...]]]]

However, no direct feature checking (as was supposed in (Reuland, 2001)) occurs here; the problem is that in (Chomsky, 2001, 2005) and PT's systems, only unvalued features can serve as agreement probes, and the unvalued features of *zich* clearly cannot find any value this way. This problem is solved if we assume that for independent reasons, the valued features of the antecedent and the unvalued features of the anaphor become members of the same feature "chain", thus getting to share the same value. (Reuland, 2005) proposes that such a dependency involves the heads Tns and v^* , which "mediate" between the antecedent and the SE anaphor. This is how it happens:

- An Agree relation is established between the instances of T feature on Tns, EA and v^* (the process assumed by (Pesetsky & Torrego, 2004b), demonstrated in (19) on page 32).
- An Agree relation is established between v^* and the object SE anaphor, resulting in movement of SE to the edge of v^* . This agreement process normally serves to value the unvalued uninterpretable ϕ -features on v^* . However, in case of a SE anaphor, the ϕ -features on the object are unvalued themselves, so ϕ -features on v^* still remain unvalued. Note that the agreement operation itself is still licit, since in PT's (Pesetsky & Torrego, 2004b) framework, nothing prevents an Agree relation between two unvalued features.
- In a normal case, the unvalued ϕ -features on Tns would probe for the valued ϕ -features on EA, resulting in the valuation of the ϕ -features on Tns. However, in the present case, this would result in a derivation crash, because the ϕ -features on v^* (and the object SE) would be still left unvalued. To save the situation, the T-dependency between Tns, EA and v^* is extended to a ϕ -feature dependency.
- As a result, we have an agreement chain of ϕ -feature instances on the following elements (Reuland's (7)):

(54) [Tns_{u ϕ} SE_{u ϕ} [EA_{val ϕ} [v*_{u ϕ} [V (SE_{u ϕ})...]]]]

In this configuration, the ϕ -features on EA provide values for all the other ϕ -features in the configuration. Consequently, the SE anaphor gets these values from the subject DP; needless to say, such an interpretation is unavailable for a pronoun in the same position, since it comes out of the lexicon with valued ϕ -features. To conclude, there is a syntactic dependency between the SE anaphor and its antecedent, which is interpreted as a binding relation.

Finally, (Reuland, 2005) notes that in his system, the antecedent of a SE anaphor need not c-command it: "If binding is encoded as Agree, it is expected that the requirement that the binder c-command the bindee

is derivative of the requirement that the SE-anaphor be a target for a (c-commanding) probe. Thus, one should find cases where the bindee can be probed by a relevant head, but its eventual antecedent does not c-command it". (p. 511–512). Indeed, such cases are found in Norwegian and Icelandic.

2.2.5 Summary

In this section I have presented the development of views on (syntactic) binding. The canonical BT treated binding as a uniform phenomenon, a relation available to several types of nominal elements, with some syntactic constraint to rule out ill-formed configurations. I have demonstrated that in the version of the BT developed in (Reinhart, 1983, et seq.) and (Reuland, 2001, 2005), which I adopt, binding as a uniform syntactic phenomenon ceases to exist. The empirical facts earlier covered by the binding principles A, B and C are now governed by distinct mechanisms:

1. Binding in the logical syntax sense (which can be understood as binding of a variable by an operator, or in Reinhart's (2000) sense, see definition in (27) on page 39).
2. Covaluation – free assignment of the same value to different DPs.
3. Reflexive-marking of reflexive predicates.
4. Formation of chains/agreement relations between DPs.

The first of these mechanisms belongs to semantics, the second one is implemented at the discourse level. Thus, only the last two types of dependencies are encoded in syntax proper. In much work on tense, sequence of tense relations are viewed as logical binding, and many authors derive it using semantic mechanisms only. I will opt for a different strategy and will show that relations between temporal morphemes can be viewed as relations of the fourth type. To do so, I will use the agreement strategy based on (Pesetsky & Torrego, 2004b) and (Reuland, 2005).

Chapter 3

Sequence of Tense

This chapter presents the problem of Sequence of Tense (SOT), that is, generally speaking, the interpretation of tenses in embedded clauses. In section 3.1, I introduce the relevant data: the interpretation of present and past tenses in complement clauses as opposed to adjunct and relative clauses in languages like English, and the crosslinguistic variation in temporal interpretations. In section 3.2, I discuss several theories that aim to explain how the interpretations of embedded tenses are derived; I analyse in some detail the approaches of Enç (1987, 2004), Abusch (1988, et seq.), Schlenker (1999, 2003, 2004), Babyonyshev & Matushansky (2006), Kratzer (1998, 2006), Stechow (2003a, 2003b), Kusumoto (1999), while briefly mentioning several others.

3.1 An overview of the data

3.1.1 SOT: The problem

The term “sequence of tense” (SOT) refers to the behaviour of tenses in embedded clauses, most notably, in cases when a complement clause is embedded under a past predicate. In languages like English, past tense embedded under a past matrix verb can render two types of interpretation: a “past-shifted reading” and a “simultaneous reading” (the terms were introduced by (Enç, 1987)). Examples (1a) and (1b) illustrate these two readings:

- (1) a. John said that Mary won the race.
- b. John said that Mary was pregnant.

The sentence (1a) can only mean that according to John, Mary won the race at some time previous to his saying so. In other words, the content

of John's speech could have been "Mary won the race!" This is the past-shifted reading: the event described in the embedded clause is ordered as preceding the matrix event.

The sentence (1b), which contains a stative verb in the embedded clause, also allows for this kind of reading, although it is considered less salient. But most importantly, it can mean that John's words were something like "Mary is pregnant now!" – that is, the event of Mary's (alleged) pregnancy was simultaneous with John's speaking about it. This is the reading that the SOT theory has been mostly concerned with.¹

A traditional view on past tense in formal semantics, which stems from (Prior, 1967), is that it is a sentential operator that shifts the time of the event back with respect to some evaluation point. In case of past embedded under past, only the matrix past will be a possible evaluation point for it. Thus, only the past-shifted reading is predicted for the embedded past. The existence of the simultaneous reading is a puzzle.

However, even if we argue that in (1b), the evaluation point for the embedded tense can be the speech time, it is still not the end of the story. The event described by the embedded verb does indeed precede the moment of speech; however, if it were evaluated with respect to speech time only, then we would also expect a "forward-shifted" (a term by (Enç, 1987)) reading for the embedded verb: that is, a reading under which the time of Mary's pregnancy is located between John's talking about it and the actual utterance time. This reading is unavailable – the sentence (1b) cannot mean 'John said that Mary would be pregnant'. Similarly, (2) is incoherent:

(2) # John said in 1999 that Mary was pregnant in 2000.

This is unexpected if we assume that the embedded past is just a regular past tense. There has to be some kind of relation established between the matrix and the embedded tense. In most accounts of SOT it is assumed that the interpretation of the embedded past tense is dependent on the matrix tense, and it is interpreted either as past with respect to the matrix tense (past-shifted reading), or as present (simultaneous reading). The derivation of the latter reading, and the precise nature of the relation between the matrix and the embedded tense, is the major problem of the SOT theory, and the phenomenon of embedded past tense interpreted as simultaneous to the matrix past tense is often called "sequence of tense" in the narrow sense. As will be shown in greater detail below, two ways of solving this problem have been proposed. One is to assume that the embedded past tense is indeed a past tense, but it is anaphoric to the matrix past,

¹An independent issue is exactly why only stative verbs allow for the simultaneous reading. Most theories of SOT leave this problem aside and simply deal with stative verbs all the time. I will do the same here, and will not attempt to build a theory that would explain this phenomenon based on aspectual properties of verbs. Let me just point out a parallel with the Russian perfective/imperfective contrast, noted in section 1.4.1: only imperfective verbs can give rise to simultaneous readings in this language.

or bound by it. The other is to consider the embedded past a “vacuous” or “null” tense, which has the appearance of past due to a semantically blind process. In this dissertation, I will present evidence in favour of the first view, and I will argue against the second.

Following (Ogihara, 1989, 1995, 1996) and other authors, I will call the type of temporal interpretation when a tense is valued with respect to speech time the *absolute*, or *independent* temporal interpretation, and the interpretation of an (embedded) tense with respect to the time of the event described by the matrix verb – the *relative*, or *dependent* temporal interpretation (however, I am not committing myself to the way Ogihara uses these terms – namely, to denote two types of tense morphemes).

Not only verbs, but also NPs that refer to a past event can trigger SOT phenomena (the appearance of simultaneous past) in their complements. This is demonstrated in the following examples from (Abusch, 1988) (her (28) and (29) on p. 10):

- (3) a. Mary’s announcement that she was pregnant is irrelevant now.
 b. Everybody knows John’s earlier claim that Mary would win the prize.

As for present embedded under past in English, it behaves differently from past embedded under past. This is also a rather surprising fact for the theory of tenses: if we allow past to be interpreted as past with respect to the matrix reference point and not the moment of speech, it should be in principle possible to interpret present in a similar way. However, this is not the case. Consider (4):

- (4) John said that Mary is pregnant.

As (Comrie, 1986) puts it, the claim about Mary’s pregnancy should have “continuing applicability” for the sentence to be true.² (Enç, 1987) terms this reading “a double-access reading”. Thus, the present tense embedded under past in English must always refer to the moment of speech.

²The exact properties of this reading are not easy to pinpoint. Of course, John might have been mistaken in his claim about Mary’s state, so it is obviously wrong to say that Mary has to be pregnant at the present moment for the sentence to be true. As Abusch (1997) points out, a more accurate explanation is that the available “symptoms” that led John to his belief (for example, Mary’s big belly) have to be still present at the moment of utterance. The speaker could therefore question John’s statement by saying something like (i) (Abusch’s (69) on p. 40):

- (i) John said two weeks ago that Mary is pregnant but actually she has just been overeating for the last three months.

Importantly, though, John himself was not making any predictions about Mary’s state at the present moment; the present tense signals that it is the speaker who interprets the statement as still being valid.

3.1.2 “Truly vacuous” past

It has been argued that there exist cases when the morphologically past verb in an embedded clause does not refer to any past moment at all. This is demonstrated by the examples in (5), which have become an important argument in the SOT theory:

- (5) a. John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together (Abusch, 1988, p. 2, (6)), originally from (Kamp & Rohrer, 1984).
- b. We decided to tell the prosecutor tomorrow that we were talking to him reluctantly (Enç, 2004, p. 205, (6)).

In both sentences in (5), the deepest embedded verb – *were* – refers to a moment that does not precede either the utterance time or indeed any other evaluation point introduced in the sentence. Rather, it is posterior to the utterance time. This happens because the second embedding (containing *would* or infinitive) refers to a moment posterior to the matrix time, and the deepest embedded verb is interpreted as denoting a time simultaneous to that moment. Surprisingly, the verb still carries past morphology. These examples, which I will informally term “last meal” examples, have been proposed as a counterargument to all theories which claim that the embedded past in SOT configurations is interpreted with respect to utterance time. I will come back to this discussion in subsequent sections.

3.1.3 Tenses in relative and adjunct clauses

As can be seen from the examples, the discussion so far only involved complement clauses. In fact, these are the only environments that exhibit SOT-effects in languages like English. Consider the interpretation of tenses in a relative clause:

- (6) We spoke to the man who was crying (Enç, 1987, p. 638, (16)).

Here the time of the event described in the embedded clause can either precede, overlap with, or follow the time of the matrix event. The latter reading is especially easy to obtain in sentences like (7) (Kusumoto, 1999, p. 95):

- (7) Hillary married a man who became the president of the U.S.

This sentence can be readily understood as stating that Hillary married the man before he became president. We can conclude that the embedded past tense is interpreted independently of the matrix tense, and denotes some moment previous to the time of speech.

As for the present tense in relative clauses embedded under matrix past, it does not have the peculiar “double-access reading” and simply refers to the speech time, without any connection to the matrix event time:

- (8) Adam gave an ice-cream cone to a boy who is sitting outside (Stowell, 1995, (6a), p. 385).

It is usually assumed that in English, relative clauses are not “SOT configurations”: no dependency between the embedded and the embedding tense can be established in such configurations. This is the conclusion reached, among others, by (Stowell, 1995), who gives a very detailed analysis of tense interpretations in different syntactic environments. The same holds for adjunct clauses; consider, for example, the temporal ordering in (9):

- (9) a. I spoke to John before he left the office.
b. I spoke to John after he left the office.

In both (9a) and (9b), past tense is used in the embedded clause; it can be interpreted as either preceding or following the time of the matrix predicate, depending on the semantics of the complementizer.

A remark is due here about the interpretation of past tense embedded under past in relative clauses. As noted above, the embedded past in (6) can be interpreted as preceding, simultaneous with, or posterior to the matrix past. From this we conclude that independent interpretation of the embedded tense is possible here – that is, the embedded tense is interpreted relative to the utterance time. However, nothing prevents us from saying that the dependent interpretation is also possible. It might be that the embedded tense can also be interpreted with respect to the matrix time – as past-shifted or simultaneous; however, these readings are simply indistinguishable from the ones that are yielded by the independent interpretation mechanism. This point is made in e.g. (Enç, 1987; Ogihara, 1996). (A parallel can be drawn with binding and covaluation of pronouns: the same interpretation can be derived either via the structural relation of binding or “accidentally” because the two elements pick out the same value.) I will return to this problem in sections 4.2, 4.3.3.

To sum up, SOT phenomena are present in some embedded contexts, but not in others. There are two ways of distinguishing the SOT and non-SOT environments. One, which I have used in the discussion above, is structural: SOT phenomena are observed in complement clauses but not in other types of embeddings. The other is semantic: SOT phenomena depend on the type of the matrix verb. According to the semantic view, only attitude verbs like *believe*, *know*, or *hear* (or, more generally, intensional predicates) trigger an SOT interpretation in a language like English. (Enç, 1987, 2004, et seq.) defends the structural approach; similarly, (Ogihara, 1989, et seq.) argues that SOT interpretations are not dependent on attitude semantics. The second view is taken by (Abusch, 1988, 1997; Stechow,

2003a, 2003b, 2004; Kratzer, 1998, a.o.), who present some arguments to show that SOT phenomena are apparently manifested in intensional contexts which are not complements.

In this thesis, I will take an intermediate position. I will argue that syntactic structure plays a role in establishing temporal interpretation. However, it is impossible not to resort to attitude semantics for the explanation of some temporal phenomena, such as the non-existence of “forward-shifted” readings. Note that this situation is by no means unique in language. Thus, the processes of anaphora resolution are governed by distinct modules of the language system: syntax proper, logical syntax, and discourse (Reuland, 2001; Koornneef, Wijnen, & Reuland, 2006). I propose to treat temporal interpretations in the same spirit.

3.1.4 Tenses embedded under future

So far, only embeddings under past matrix verbs have been mentioned. Still, future embeddings have also been examined in the theory of SOT. As regards complement clauses, the facts, as presented e.g. in (Stowell, 1995), are the following: when a complement clause is embedded under a future verb, both present and past tenses in the embedded clause can be interpreted as relative tenses – that is, past precedes the matrix event time, and present overlaps with it:

- (10) a. John will say in 2010 that Mary was pregnant in 2009.
 b. In 2010, John will say that Mary is pregnant.

With relative clauses things are more complicated. (Stowell, 1995) concludes that tenses in relative clauses embedded under future can only be interpreted as absolute, just as in relative clauses embedded under past:

- (11) a. Adam will give an ice-cream cone to a boy who was sitting outside.
 b. Adam will give an ice-cream cone to a boy who is sitting outside.
 (Stowell, 1995, (7a,b), p. 385)

According to Stowell (1995), sentences (11a) and (11b) can only mean that the boy was sitting outside before the moment of speech or is there at the moment of speech, respectively. The time of sitting cannot be interpreted relative to the moment of Adam’s activity.

However, several authors claim that at least some instances of past and present in relative clauses in the scope of a future tense can be interpreted relative to the matrix tense; the relevant examples are given in (12):

- (12) a. John will buy a fish that is alive (Ogihara, 1999a, ((13), p. 229)).
 b. I will charge you whatever time it took (Heim, 1994, p. 158, (42)).

- c. At the end of next term, I will give automatic A's to all students who turned in their term papers on time (Abusch, 1998, p. 13, (1)).

Sentence (12a) means that the fish John is intending to buy will be still alive at the time of his buying it. In (12b), the job for which the speaker is going to charge does not have to be commenced already. Similarly, according to Abusch (1998), sentence (12c) can be uttered when no term papers have been turned in yet. These examples suggest a special analysis for future embeddings, although few authors (apart from Abusch, 1998, and some others) pay much attention to these cases. I turn to this problem in section 4.3.4.

3.1.5 SOT facts cross-linguistically

There are quite a few languages which behave in a way similar to English with respect to SOT. Dutch, for example, displays the same pattern. For a simultaneous interpretation of the matrix and complement clauses, embedded past is used:

- (13) Jan zei dat Marie zwanger was.
Jan said that Marie pregnant was

Present embedded under past has a double-access interpretation:

- (14) Jan zei dat Marie zwanger is.
Jan said that Marie pregnant is

And to unambiguously express a past-shifted interpretation, Dutch uses pluperfect:

- (15) Jan zei dat Marie zwanger was geweest.
Jan said that Marie pregnant was been
'Jan said that Marie had been pregnant'

The same can be said about Romance languages such as Spanish, French and Italian. For example, in Spanish, subjunctive obeys SOT, as can be seen in the examples (16) (from (Comrie, 1986, p. 270, (11)-(12)), glosses mine):

- (16) a. Eduardo me dice que lo haga.
Eduardo me tells that it do-SUBJ.PRES
'Eduardo tells me to do it'.
b. Eduardo me dijo que lo hiciera.
Eduardo me told that it do-SUBJ.PAST
'Eduardo told me to do it'.

The languages that display this pattern of tense interpretation in embedded clauses have been termed “SOT languages”. So, SOT languages are those languages where in a complement clause embedded under a past verb, past tense is used to express simultaneity, and the use of present can only yield a double-access reading.

However, not all languages follow this pattern. A different type of tense behaviour is represented by the so-called “non-SOT languages”. In languages of this type, past tense embedded under past is interpreted as prior to the embedded past (in other words, it normally allows only for a past-shifted reading). To get a meaning of simultaneity between the matrix and the embedded predicate, present is used in the embedded clause.

One well-known example of a non-SOT language is Russian. In Russian complement clauses, past embedded under past is usually interpreted as denoting precedence, and present embedded under past is used to obtain a simultaneous reading. Thus, Russian present does not have to always refer to speech time – it can be evaluated as present with respect to some point or interval in the past.

- (17) a. Ivan skazal, čto Maša bolela.
 Ivan said that Masha ailed
 ‘Ivan said that Masha had been ill’
 b. Ivan skazal, čto Maša boleet.
 Ivan said that Masha ails
 ‘Ivan said that Masha was ill (at the time of saying)’

However, a claim that in Russian past embedded under past *always* denotes precedence in complement clauses would be too strong. It has been noted that there are some contexts where Russian displays a simultaneous reading of past under past. Often this happens if the matrix verb is a perception verb. This phenomenon is discussed, for example, by (Altshuler, 2004), as illustrated in (18) (his (1) and (2)):

- (18) Dina videla, čto/kak voda lilas’/l’jětsja iz vedra.
 Dina saw that/how water poured/pours from basket
 ‘Dina saw that/how the water was pouring from the basket’

I will focus on these data in section 4.1.2 and show that they present an important piece of evidence in the theory of SOT. Namely, I will show that embedded past can be interpreted as simultaneous to the matrix past and still retain a connection with the utterance time. In particular, this shows that the problem of the absence of forward-shifted readings cannot be solved by introducing a “vacuous” tense, since forward-shifted interpretation of embedded past is unavailable in Russian, too, as demonstrated in (19).

- (19) #V 1999 godu Ivan skazal, čto Maša byla beremenna v 2000
 in 1999 year Ivan said that Masha was pregnant in 2000
 godu.
 year
 #‘Ivan said in 1999 that Masha was pregnant in 2000’

This example also serves to show that although Russian (along with other languages) has been termed a “non-SOT language”, this does not mean that it does not display any interaction between tenses in a sentence. In other words, the choice of tense forms in Russian embedded clauses is by no means arbitrary. Rules that govern the behaviour of tenses in Russian complement clauses do not apply to adjunct and relative clauses, as can be shown by the examples in (20):

- (20) a. Ivan uvidel devočku, kotoraja sidela na skamejke.
 Ivan saw girl which sat on bench
 ‘Ivan saw a girl who was sitting on a bench’
 b. Ivan uvidel devočku, kotoraja (sejčas) sidit na skamejke.
 Ivan saw girl who (now) sits on bench
 ‘Ivan saw a girl who is sitting on a bench (now)’

In (20a) and (20b), tenses in relative clauses have the same interpretations as their English counterparts. Past embedded under past in a relative clause is interpreted with respect to the utterance time; it can have a past-shifted, a simultaneous and a forward-shifted interpretation. Present embedded under past denotes the speech time and cannot be interpreted as present with respect to the matrix time (that is, (20b) cannot mean that the girl was sitting on the bench at the time of Ivan’s seeing her).

The Russian relative clause data are in contrast with the data from Japanese – another well-studied example of a “non-SOT” language. In Japanese, tenses behave differently from English not only in complement clauses but in relative and adjunct clauses as well. The use of tenses in Japanese complement clauses is demonstrated by the Japanese examples from (Ogihara, 1996) (his (2), chapter 3, p. 69):

- (21) a. Taro-wa [_SHanako-ga byooki-da]-to it-ta.
 Taro-TOP Hanako-NOM be-sick-PRES that say-PAST
 ‘Taro said that Hanako was sick [at that time]’
 b. Taro-wa [_SHanako-ga byooki-dat-ta]-to it-ta.
 Taro-TOP Hanako-NOM be-sick-PAST that say-PAST
 ‘Taro said that Hanako had been sick’

Sentences (21a) and (21b) demonstrate the same choice of tenses as observed in Russian. However, in relative clause in (22a) and in adjunct clauses in (22b) and (22c), embedded tenses are interpreted with respect to the time of the matrix clause event and not the speech time:

- (22) a. Taroo-wa [nai-te i-ru otoko]-o mi-ta.
 Taro-TOP cry-PROG-PRES man-ACC see-PAST
 ‘Taro saw a man who was crying [at the time of the meeting].’
 (Ogihara, 1996, (5), chapter 5, p. 153)
- b. Taroo-wa [Hanako-ni au mae-ni] denwa-o
 Taro-TOP Hanako-DAT meet-PRES before phone-ACC
 si-ta.
 do-PAST
 ‘Taro called Hanako before he saw her.’ (Ogihara, 1996, (55a),
 chapter 5, p. 181)
- c. Taroo-wa [Hanako-ni at-ta ato-de] denwa-o
 Taro-TOP Hanako-DAT meet-PAST after phone-ACC
 si-ta.
 do-PAST
 ‘Taro called Hanako after he saw her.’ (Ogihara, 1996, (55b),
 chapter 5, p. 181)

For this reason, some researchers further divide the class of non-SOT languages into the “Russian type” and the “Japanese type”. However, (Kusumoto, 1999) shows convincingly that the behaviour of tenses in Japanese relative and adjunct clauses can be explained independently by the structural properties of the clauses: Japanese relative clauses are shown to be structurally analogous to English participles, and adverbial clauses are analysed as having the same structure as relatives. I will outline Kusumoto’s account in more detail in section 3.2.2; it is important for me because it demonstrates that there are no “Japanese-type” non-SOT languages, and so non-SOT phenomena can be treated uniformly. Hence, my analysis of Russian could in principle cover all non-SOT languages; however, detailed testing of this prediction is outside the scope of the present work.

3.2 Theories of SOT

The problem of Sequence of Tense, as well as the syntax and semantics of tense in general, has received considerable attention in the literature over the past decades. Consequently, the amount of literature on the topic is quite daunting. I will not try to review all existing SOT theories in detail; instead, I will give a general outline of the debate and then concentrate on the findings that I will rely on in constructing my own account. I will also focus on how different theories fare with respect to Russian SOT facts.

The plan for this section is the following. First, I present two approaches to the nature of tenses: tenses as sentential operators and as referential expressions. The first (Prior, 1967, a.o.) treats tenses as operators which have scope over sentences. SOT phenomena are a problem

for this account, and a special SOT rule is needed to deal with them. The second (starting from Partee, 1973) takes tenses to be referential expressions, sharing important properties with pronouns. Thus, it is assumed that like pronouns, tenses can be used both deictically and anaphorically. This approach provides a natural solution for the SOT facts: namely, that the embedded tense is bound by the higher tense. However, the exact implementation of this intuition is far from obvious. In the first place, it is debated what is the binding mechanism in question. Enç (1987) treats temporal binding as syntactic binding in the spirit of (Chomsky, 1981). Abusch (1989, et seq.) argues against putting SOT into syntax proper, and proposes a mechanism of logical syntax binding for embedded tenses. The latter path is also followed by (Kratzer, 1998; Stechow, 2002, et seq.; Kusumoto, 1999). The referential tense theories mentioned above explore the analogy between tenses and anaphoric expressions; in contrast, Schlenker (1999, et seq.) uses the parallel between indexicals and present tenses, incorporating them into his typology of shiftable/non-shiftable indexicals.

In my own account, I will also pursue the pronominal analogy; however, I will argue against putting temporal binding in logical syntax, and I will adopt Enç's idea that temporal morphemes can be bound in syntax proper, although I will not use the classical BT binding mechanism that she resorts to. I will also incorporate Schlenker's ideas into my proposal, and argue that "indexical shifting" is indeed a suitable mechanism for deriving the interpretation of Russian present-under-past.

3.2.1 The nature of tenses

The starting point of any research on tenses is to establish what the nature of tenses is and how temporal interpretations are computed. As was already noted in section 3.1.1, one approach, traditionally used in formal semantic frameworks such as Montague's (1973) PTQ, views tenses as sentential operators. According to this view, "[e]ach expression is interpreted with respect to a world and a time (called indices), and a tense operator shifts the time index away from the original index" (Ogihara, 2003). An example of a sentence interpretation is given in (23) (Ogihara, 2003, (1)), where **P** is a past tense operator:

- (23) a. Every student left.
 b. $\mathbf{P}[\forall x[\text{student}(x) \rightarrow \text{leaves}(x)]]$
 c. There is some past time t such that every student x at t leaves at t .

Straightforwardly, this approach can only give us a past-shifted interpretation of a past tense embedded under past. This is so because a sentence like (24a) (Enç, 1987, p. 635, (5)) has to be analysed as (24b), where the matrix past operator takes scope over the whole sentence, including the embedded past operator, so the latter can only shift the temporal index of its

clause with respect to the matrix time. This representation only yields the past-shifted reading of this sentence (with Mary's alleged pregnancy preceding John's hearing about it), but not the simultaneous reading (where the two eventualities overlap). A solution to this problem is to introduce an "SOT rule", which makes sure that the embedded tense is not interpreted as past. In fact, it is not considered past at all, but rather present tense or "zero tense" which surfaces as past due to the application of the SOT rule; thus, the simultaneous reading for (24a) is represented by (24c) (Enç, 1987, p. 635, (9)).

- (24) a. John heard that Mary was pregnant.
 b. PAST [John hear [PAST [Mary be pregnant]]]
 c. PAST [John hear [PRES [Mary be pregnant]]]

There are two ways to implement the SOT rule. In the framework of (Ladusaw, 1977; Comrie, 1986), the rule replaces present tense with past.³ As Enç (1987) points out, this must be a late morphological rule operating at PF, since LF must not see the replacement. Similarly, if a special zero tense is introduced, then the grammar provides a rule for spelling it out as past in certain configurations.

In the work of Ogihara (1989, et seq.), another version of the SOT rule is used: it is a rule that optionally deletes past tense at LF when it is in the (local) scope of another past tense to yield an "empty" tense that is interpreted as simultaneous to the higher tense. Consequently, the rule belongs to semantics and not to syntax. How it works can be seen in (25) (Ogihara, 1996, (4), chapter 4, p. 103):

- (25) a. (*Surface sentence*) John said that Mary was pregnant.
 b. (*Underlying structure*) John Past say that Mary Past be pregnant.
 c. (*LF after application of SOT rule*) John Past say that Mary \emptyset be pregnant.

Ogihara's SOT rule does not distinguish between clause types (complement clauses vs. adjunct and relative clauses) or between types of embedding verbs (attitude verbs vs. non-attitude verbs), so it yields dependent tense readings in all types of embedded clauses. Indeed, Ogihara assumes that in English relative clauses tenses are also dependent on the matrix tense, but this dependency is not visible since an independent reading of tense is also available in this type of clauses.

For Ogihara, the difference between SOT and non-SOT languages is a matter of parametric variation: the SOT rule is simply absent in non-SOT languages. He illustrates his point with examples from Japanese, where

³Comrie (1986) does not use the operator theory of tenses; however, his work should be mentioned since it contains one of the most well-known formulations of the SOT rule.

a simultaneous reading of a verb embedded under past is expressed by present in all types of embedded clauses (see examples (21)-(22) in section 3.1.5). To account for the fact that in English, present tense can also be embedded under past yielding a double-access reading, Ogihara suggests that English and Japanese present tenses are different in nature. In (Ogihara, 1999a, 2003) he argues that past tense both in English and Japanese can be sometimes an absolute tense and sometimes a relative tense; however, present tense in all languages can only be an absolute tense. What looks like a relative present tense in Japanese is in fact a tenseless form, which behaves very much like the English present participle. English lacks such a form, and that is why it needs a SOT rule. While this account can distinguish between English and Japanese, Ogihara himself admits that it does not work for languages such as Russian and Polish, where present tense is interpreted as absolute in relative and adjunct clauses but not in complement clauses. Thus, it is hard to maintain this version of SOT rule if we want to cover both Japanese and Russian data.

An alternative view on tenses is proposed in (Partee, 1973), who proposes to treat tenses as referential expressions, similar to pronouns. She notes several parallels between the interpretation of tenses and pronouns. Thus, tenses, as well as pronouns, can be interpreted deictically or anaphorically. For example, the sentence (26) (Partee, 1973, p. 602, (3)), uttered by the speaker when she has just left home, is analysed as an example of a deictic past: its tense refers to “a definite interval whose identity is generally clear from the extralinguistic context” (Partee, 1973, p. 603). The operator theory has a problem with this example, since “such a sentence clearly does not mean either that there exists some time in the past at which I did not turn off the stove or that there exists no time in the past at which I turned off the stove” (Partee, 1973, p. 602). In other words, just a shifting of the temporal index is not enough for the interpretation of tenses: there is a particular interval that a past tense refers to.

(26) I didn't turn off the stove.

Partee (1973) also gives examples of a temporal interpretation which is similar to the anaphoric interpretation of pronouns, as in (27): here, “the time is specified in one clause and the tense of a subsequent clause refers to the same time” (Partee, 1973, p. 605). (28) is given as an example of a bound variable use of tense.

(27) Sheila had a party last Friday and Sam got drunk. (Partee, 1973, p. 605, (10))

(28) If Susan comes in, John will leave immediately. (Partee, 1973, p. 605, (14))

This picture is of course somewhat simplified. (Partee, 1984) argues that tenses do not ‘refer’ to times in the same sense as pronouns refer to entities.

For example, in a text like (29), the events described by predicates *got up*, *went to the window*, etc, move the narrative forward by shifting the reference time of the past tense. However, the simple system outlined in (Partee, 1973) only allows tenses to refer back to some interval of time already mentioned in the discourse, not to a time that would follow some beforementioned time. To incorporate such finer temporal relations into the theory, Partee (1984) proposes an account of tenses based on the DRT of (Kamp, 1981) and on the temporal theory of (Hinrichs, 1986).

- (29) John got up, went to the window, and raised the blind. It was light out. He pulled the blind down and went back to bed. He wasn't ready to face the day. He was too depressed. (Partee, 1984, p. 253, (14))

In any case, the pronominal theory of tense proves useful for the theories of SOT if we assume that tenses, like pronominal elements, can be bound. From this point of view, the dependency between the higher and the lower tense in a sentence like (24a) on page 60 on a simultaneous reading can be derived via a binding relation, similar to a relation between an anaphor or pronominal and its antecedent. This is the line pursued by many authors, among them (Enç, 1987, 2004; Kratzer, 1998; Stechow, 2003a, 2003b). However, they take different stands with respect to the nature of the binding relation between tenses.

The most straightforward way to implement the binding intuition is to treat temporal morphemes as pronominal elements that can be syntactically bound in the sense of the pronominal Binding Theory. As was discussed in chapter 2, the classical BT of (Chomsky, 1981) treated the binding of pronominals and anaphors as a syntactic phenomenon. The prerequisites of binding were coindexation and a c-command relation between the binder and the bindee, and the theory imposed syntactic constraints on the configurations in which binding was grammatical, summarized in the binding principles, or conditions (see (25) on page 38). Enç (1987) uses this model to account for the SOT phenomena: roughly, in her account tenses can be bound by higher tenses when they are coindexed with them and c-commanded by them, and several constraints are introduced to make sure that binding is licit in complement clauses but not in other types of embeddings.

On the other hand, many authors argue for what can be called a semantic theory of SOT. Under this view, temporal morphemes are treated as variables which are bound at the level of logical syntax. Since logical syntax binding does not imply any constraints on the configurations that the binder and the bindee are found in (as long as they are in a c-command relation), this theory is in need of special conditions for the contexts in which the bound interpretation of tenses is licit. These contexts are usually taken to be attitude contexts, as in (Kratzer, 1998; Stechow, 2003a, 2003b; Kusumoto, 1999, and others). In what follows, I will discuss these approaches

and develop some of the ideas expressed by the authors; however, I will argue for a syntactic binding procedure for temporal morphemes.

3.2.2 Accounts of SOT

In this section I discuss several theories of SOT. I will concentrate on those of them which contain ideas that I will rely on for my own proposal, as well as on those which deal specifically with “non-SOT” languages like Russian.

Enç (1987, 2004)

Enç (1987, 2004) proposes an account that treats tenses as referential expressions that can be bound. Enç (1987) assumes that tense is in Infl and that it has to bear an index, just like all referential expressions do in the BT of (Chomsky, 1981). Tenses denote intervals, and “[t]he interval assigned to tense according to its index is the temporal argument of the verb” (Enç, 1987, p. 640). Further, Enç argues that tenses (PAST and PRESENT) are relational notions, always determined with respect to some evaluation point/interval. This evaluation point is provided by the local Comp, which can optionally carry a temporal index. In matrix clauses, Comp denotes the speech time, and the temporal interpretation of the verb is computed with respect to this interval (precedence for PAST, simultaneity for PRESENT). Embedded tenses can be linked to speech time via higher (embedding) tenses. The relative nature of tense is captured in the Anchoring Principle: “Each tense must be anchored” (Enç, 1987, p. 642, (26)), that is, it must eventually be evaluated with respect to speech time. The rules governing anchoring, which Enç models along the lines of the binding principles, are given in 30 (Enç, 1987, p. 643, (27)):

(30) *Anchoring Conditions*

- a. Tense is anchored if it is bound in its governing category, or if its local Comp is anchored. Otherwise, it is unanchored.
- b. If Comp has a governing category, it is anchored if and only if it is bound within its governing category.
- c. If Comp does not have a governing category, it is anchored if and only if it denotes the speech time.

Thus, there are two options for anchoring of tenses: either a tense is related to its local Comp, which must be either bound or denote the speech time; or it is itself bound. The notions of binding and government category are taken from the (classical) BT (see the definitions in (23) and (25), section 2.2.1), which means that a tense can be bound by a c-commanding tense with the same index. This results in a dependent tense reading, such as the simultaneous reading of past in SOT configurations.

Without going into the technical details of the analysis, below I show how tenses in different syntactic configurations get their interpretation in Enç's model.

1. Matrix clauses. In matrix clauses, Comp always denotes the speech time, so it is anchored, which means its tense is anchored as well.
2. Relative clauses. There are two ways to interpret the embedded tense in relative clauses. First, the governing category of the embedded tense is the matrix clause, so it can be bound by the matrix tense.⁴ This gives us a dependent temporal reading, which, as argued also by (Ogihara, 1996), is possible for tenses in relative clauses. Second, the tense can be anchored via its Comp if it is anchored. Comp in relative clauses has no governing category, so it can only denote the speech time; thus, if the tense in a relative clause is anchored via Comp, it is evaluated with respect to speech time, which gives us the independent interpretation of the embedded tense.
3. Complement clauses. First, Enç analyses past embedded under past. In this case, again, the matrix tense has a governing category, so it can be governed either via Comp or via the higher tense. The Comp of the complement clauses has a governing category, so it is bound by the higher tense, which means that it denotes the same time interval. Thus, if the complement past tense is anchored via its Comp, it will be interpreted as shifted with respect to the matrix tense. The second option is binding of the embedded tense directly by the matrix tense, which yields the simultaneous reading.

Enç suggests that for present tense, Comp always denotes the same interval as its tense. This by itself does not mean that present has to be inherently connected with the speech time. Indeed, Enç notes that in languages like Russian present tense can be interpreted independent of the utterance time, but this is impossible in English; thus, languages differ as to whether present tense denotes the speech time. In English, both present tense and its Comp denote the speech time in matrix as well as in embedded clauses. This creates a problem for the treatment of present tense embedded under past in complement clauses, which in Enç's system has to be bound by the matrix past. To account for this, Enç introduces a reindexing rule that changes the index of the English present tense and its Comp to speech time at LF. Further, Enç (1987) suggests that in non-SOT languages like Russian or Japanese Comp always carries a temporal index, so tenses have to be anchored via their Comps and cannot be bound by higher tenses. This means that embedded past in these languages will always have a past-shifted reading.

⁴For the sake of brevity, I am not going into the details of how the governing category is computed; the reader is referred to (Enç, 1987).

An important property of Enç's (1987) theory is that all tenses, both matrix and embedded, eventually get anchored to speech time. This means that past always precedes the speech time: either it is directly anchored to speech time, or it is bound by another past, or shifted with respect to another past, but in any case it ends up as a "real" past. This treatment of past gave rise to criticism in (Abusch, 1988). Abusch provides examples where a past tense is embedded under "would" or an infinitive which is in turn embedded under past, such as the "last meal" sentence (5a) on page 52, repeated here as (31), where the deepest embedded past does not refer to any past moment:

- (31) John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together.

In (Enç 2004), Enç modifies her analysis so as to incorporate these cases as well. She now distinguishes between two indices on each I, one for evaluation time (which is Enç's term for Reichenbachian reference time) and one for event time, instead of anchoring alternatively Comp or Tense. She also eliminates the notion of governing category and uses the notion of "local Tense" instead. However, the general principle of her account remains, and she argues again that tenses cannot be bound in non-SOT languages.⁵

To sum up, Enç's theory appears to fare well with much empirical data, however it does have problems. First, it cannot be implemented without resorting to indices, which have been largely eliminated from the current syntactic theory. Second, it is problematic for non-SOT languages like Russian. As mentioned above, Enç's account only predicts a past-shifted reading for the Russian past-under-past; however, as noted in section 3.1.5 above, simultaneous readings are also available for it under special circumstances. The differences between Russian and Japanese also remain unexplained: in Enç's model, present tense in non-SOT languages must be able to behave as a relative tense in all configurations, including relative clauses. This is not the case for Russian.

In my account, I will use Enç's intuition that tenses can be bound by a syntactic mechanism. However, I will propose a new implementation of this idea. To deal with Abusch's criticism, I will argue for a separate treatment of the "future-shifted" past found in "last meal" examples.

Abusch (1988 et seq.)

Abusch's (1988, et seq.) account is another theory that treats tenses as referential expressions. However, Abusch argues that SOT cannot be analysed in terms of surface syntax. She gives several arguments to support

⁵(Kondrashova, 1999, 2006) provides an account of SOT in non-SOT languages, such as English and Japanese, in the spirit of (Enç, 1987). However, in addition to temporal binding, Kondrashova also postulates a morphological SOT rule which is present in English but not in Russian and Japanese.

this view. One argument is that sometimes there is no overt T to trigger SOT – this is the case of complements of nouns such as *desire*, *claim*, etc, that have attitude semantics, like in the sentences in (32):

- (32) a. I know Mary was a strange child. But her desire to marry a man who resembled her is really bizarre. (Abusch, 1997, p. 29, (55))
 b. Mary's announcement that she was pregnant is irrelevant now.

In fact, the treatment of examples like (32b) depends on a particular analysis of NPs. If it is assumed that NPs carry a temporal feature (for instance, connected with case, as argued in (Pesetsky & Torrego 2001, 2004a, 2004b)), then they are not a problem for a syntax-based analysis. However, (32a), which involves a relative clause and not a complement clause, is still problematic. I will propose an analysis for the “future seen from the past” examples like the “last meal” examples and (32a) in section 4.1.

Abusch further argues that SOT effects are triggered by intensional contexts only. This is demonstrated by the difference in the interpretation of relative clauses embedded under an attitude verb (Abusch, 1988, p. 4, (10)):

- (33) John looked for a woman who married him.

The tense in the relative clause allows for a past-shifted or forward-shifted interpretation. In addition, the NP modified by the relative clause can receive a *de re* or *de dicto* (extensional or intensional) interpretation. The interesting fact is that the forward-shifted interpretation for the tense is unavailable if the NP is understood *de dicto* (that is, the sentence cannot be understood to mean that John looked for some non-specific woman who would later marry him). Abusch suggests that this is because in case of the *de re* reading, the NP has widest scope, and the embedded verb is not in the scope of the matrix verb at LF. But in case of a *de dicto* reading, the whole NP stays in the scope of the intensional verb. The generalization is that past tense cannot have a forward-shifted reading when it is in the scope of an intensional verb. Once the matrix verb is extensional, the problem disappears, as in the sentence (34) (Abusch, 1988, p. 5, (15)):

- (34) John bit a donkey which kicked him.

Abusch points out that these problems remain unsolved within the limits of surface syntax, that is, in an analysis like that of Enç (1987). However, examples like (33) do not necessarily prove a syntactic theory of SOT wrong; what they demonstrate is that the absence of a forward-shifted reading must have a semantic explanation: it is prohibited in the scope of intensional verbs.

The most fully-fledged version of Abusch's account is given in (Abusch, 1997). Abusch (1997) proposes two different ways of accounting for SOT

phenomena: an extensional, or independent, and an intensional theory. The independent theory is Abusch's starting point; according to it, tense is always evaluated with respect to utterance time, whether embedded or not (thus, past is always "real" past), and tenses in extensional and intensional contexts have the same meaning. In case of a simultaneous reading of an embedded tense, it is interpreted as anaphoric to the matrix tense; it can also be anaphoric to another tense in the preceding discourse.

Abusch (1997) argues that such a theory is possible, and that in this case tenses are analysed *de re*. A *de re* interpretation was introduced for NPs in (Kaplan, 1969) and (Lewis, 1979) to account for certain types of belief reports, as illustrated in the following scenario from (Quine, 1956), quoted by Abusch:

The example is of Ralph who believes of Ortcutt that he is a spy. In Quine's story Ralph has glimpsed Ortcutt in a brown hat and believes that the man glimpsed by him is a spy. On another occasion he glimpsed Ortcutt with gray hair and believes that the man he has so glimpsed is not a spy. If the object of the belief is a proposition (a set of worlds), then Ralph must be accused of having contradictory beliefs, since the proposition that Ortcutt is a spy is logically incompatible with the proposition that Ortcutt is not a spy. (Abusch, 1997, p. 5)

The *de re* theory involves the notion of an acquaintance relation, which is a certain relation that links the believer with the object of attitude. In addition, objects of attitude are not propositions, but sets of centered worlds (where a centered world is a pair of a world and its designated inhabitant, that is, the believer's notion of self). Thus, in the sentence (35), "the semantic value of the NP *Ortcutt* does not contribute to the believed centered proposition: its place is taken by an individual picked out by the acquaintance relation" (Abusch, 1997, p. 8).

(35) Ralph believes that Ortcutt is a spy.

In other words, it is not "the real" Ortcutt that Ralph's belief is about, but rather the individual in Ralph's centered belief world that Ralph is acquainted with. Thus, if the acquaintance relation was established between Ralph and Ortcutt wearing a brown hat, (35) is true (since Ralph believes that the individual in the brown hat, which in fact corresponds to Ortcutt, is a spy). But if the acquaintance relation was between Ralph and Ortcutt with gray hair, (35) is false.

Furthermore, Abusch argues that a pronoun in an intensional context is always interpreted *de re* if it is anaphoric to an NP outside this context. She goes on to argue that the same is true for tenses: whenever a tense in an intensional context is anaphoric to another tense outside it, it can be given a *de re* interpretation. She analyses an example given in (36):

- (36) The defendant Past₂ was actually at home watching ‘The Simpsons’ at the time₂ of the crime. But after hearing the testimony of the first eye-witness, the jurors clearly Past₃ believed that he Past₂ was in the laboratory building. (Abusch, 1997, p. 12)

Abusch proposes that the past tense on the verb *was*, embedded in a belief context, can be interpreted *de re*, and gives the following explanation: “When the embedded Past₂ is interpreted *de re*, it denotes the time at which the crime occurred in the real world. But this time does not figure directly in the condition on the centered belief worlds of the jurors” (p. 12). Its place is taken by a time picked out by an acquaintance relation: the jurors are indirectly acquainted with the time of the crime from the words of the first eye-witness, and they are speculating about the defendant’s whereabouts at this temporal point. Importantly, the tense of *was* is interpreted with respect to utterance time, and there is some past temporal point in the juror’s belief world that corresponds to it. The same reasoning applies to cases when the embedded past has a simultaneous interpretation: it is then interpreted as anaphoric to the matrix tense; in that case, the acquaintance relation in question is the identity to the now of the believer.

However, this theory by itself does not explain the absence of a “forward-shifted”, or “later-than-matrix”, reading. To rule out this reading, Abusch proposes that embedded tenses have to be evaluated with respect to the now of the intensional context, and that future is “not sufficiently determined” from the perspective of that point. We can only talk about the future by quantifying over future contexts using modal operators, since we cannot claim anything definite about it. Abusch formulates this as the Upper Limit Constraint (ULC) – a constraint that puts the now of the intensional predicate (the matrix eventuality time) as the “upper limit” for the interpretation of an embedded tense.

Further, Abusch shows that the *de re* theory faces several problems, the most important of which are the “last meal-sentences” like (5a), where the past tense ends up not referring to any time point preceding utterance time at all. Therefore, Abusch concludes that a *de re* theory is not viable, or at least not sufficient. What is left is an intensional theory, where the embedded tense is understood *de dicto*. The account that Abusch proposes involves the following rule: at LF, temporal relations in intensional contexts are transmitted from a higher to a lower tense via a feature-transmitting mechanism; these relations are accumulated down the tree. For past tense, the constraint is that at least one in the resulting set of relations must be the temporal precedence relation.

Thus, for an embedded past tense in an intensional contexts, two possibilities are open: either both the lower and the higher tense express the precedence relation to their local evaluation times (which gives us the past-shifted reading), or only the higher (matrix) tense expresses the precedence relation, and the embedded tense is coindexed with it, which gives

the simultaneous reading. The readings provided by the feature transmission mechanism are de dicto, and tenses end up being interpreted with respect to their local evaluation time (the belief time). In a way, this account is reminiscent of Ogihara's formulation of the SOT rule: in case of the simultaneous reading, the "pastness" of a verb embedded under past "does not count" for interpretation.

Importantly, Abusch argues that although a de re interpretation of embedded tense is not always available, its existence cannot be ruled out. She concludes that tense can in principle be interpreted both de dicto and de re.⁶ It follows that the upper limit constraint is a general constraint on tense nodes, so the belief time is the upper limit for the interpretation of its embedded time even when the latter is not actually evaluated with respect to the former (in the de re cases).

The ULC has the effect of solving another notorious problem of the SOT theory – the problem of present embedded under past, which has a double-access reading; the relevant example is repeated below as (37):

(37) John said that Mary is pregnant.

As noted in section 3.1.1, the claim about Mary's pregnancy should have "continuing applicability" (in terms of Comrie, 1986) for the sentence to be true. Abusch (1997) admits that her feature transmission mechanism fails in the case of present embeddings; thus, the present tense here has to be interpreted de re, that is, it is valued with respect to the utterance time (so it is simultaneous to it) and not the matrix/belief time. However, the ULC does not allow the time denoted by an embedded tense to completely follow the embedding time; it has to either precede or overlap with this time. Hence, the time interval that the embedded present in (37) refers to has to span two times: the time of John's saying and the utterance time.

In the following chapter, I propose that Abusch's independent theory of tense is well equipped to derive the interpretation of simultaneous past in Russian. Comparing Russian simultaneous past to English, I will show that it is indeed necessary to distinguish between the de re and de dicto interpretations of embedded tense. I will argue that in the case of Russian, the simultaneous past is indeed interpreted de re and anaphoric to the matrix past.

For SOT languages, Abusch argues that embedded tenses in intensional contexts can be interpreted both de dicto and de re. She concludes that a sentence such as *Mary believed it was raining* has two grammatical LFs, one where the embedded tense is interpreted de dicto, and one where it is interpreted de re. I agree with the intuition that the simultaneous

⁶Abusch does not specify whether the de re reading of the embedded past should only be possible when it is anaphoric to another past tense overtly present in the discourse. It does not seem necessary, though; we can imagine a de re analysis of an example like *John thought that Mary went to a private school*, where the embedded tense can denote a time in the real world when Mary was a schoolgirl.

meaning of the past can be derived in two ways. However, it is hard to see how Abusch's feature transmission analysis can be made to work for non-SOT languages. Since the ability to transmit features down the tree is taken to be a specific property of intensional predicates, such predicates in languages like Russian will have to display properties different from the English ones. This does not seem plausible, especially as other properties of intensional predicates, such as the ability to induce Upper Limit effects and *de re* interpretations, is shared by Russian intensional predicates.

Instead, I propose that the two derivations of the simultaneous reading of the English past under past are better analysed as an ambiguity between a bound reading and a covaluation reading. In the first case, the embedded tense is bound by its embedding tense. In the second case, there is no syntactic relation between the two; however, the two tenses share the same value, so the embedded tense is interpreted as anaphoric to the higher tense.

Schlenker (1999, 2003, 2004); Babyonyshev & Matushansky (2006)

(Schlenker, 2003) develops a general theory of indexicals, which in his view covers tenses as well. His approach is based on the criticism of (Kaplan, 1989), who examines an important property of indexical expressions (such as the 1st and 2nd person pronouns): they are directly referential, that is, they can only refer to the context of the actual speech act; for example, the pronoun *I* can only denote the actual speaker, but not the speaker of a reported speech situation. This idea is summarized as the fixity thesis: "The semantic value of an indexical is fixed solely by the context of the actual speech act, and cannot be affected by any logical operators" (Schlenker, 2003, p. 29, (1)). Hypothetical operators that could shift the context of evaluation for indexicals are termed 'monsters' by Kaplan.

Schlenker argues that monsters do indeed exist, and that some indexicals can be shifted. This can happen when they are in the scope of an attitude operator, as is the case with the first person pronoun in Amharic. In (38) (Schlenker's (3), p. 31), the Amharic pronoun meaning 'I' refers to John, the speaker whose discourse is reported:

- (38) Situation to be reported: John says: 'I am a hero'.
- a. *Amharic* (lit.): John_i says that I_i am a hero.
 - b. *English*: John_i says that he_i is a hero/*John_i says that I_i am a hero.

Schlenker concludes that attitude operators are monsters in Kaplan's sense, because they are "quantifiers over contexts of thought or of speech" (p. 32). An attitude operator binds a context variable, which introduces the context of the reported speech/thought act. According to the traditional view, attitude operators are quantifiers over possible worlds; in Schlenker's

theory an attitude verb introduces not only a world of the reported utterance, but also a speaker, (sometimes) a hearer, and a time of the reported utterance. An indexical in the complement of the attitude verb can depend either on the actual speech context, or on the embedded context. The Amharic sentence above can be analysed as (39) (Schlenker's (4), p. 32):

- (39) SAY_{<John, now, actually>} c_i be-a-hero (agent(c_i), time(c_i), world(c_i))

Indexicals in various languages differ in whether they can only depend on the context of the actual utterance, or also on reported contexts introduced by an attitude verbs. According to this property, all indexicals are divided into shiftable and non-shiftable. Thus, English 1st person pronouns are non-shiftable, while their Amharic counterparts are shiftable. A similar point is made for present tenses in English and Russian. Schlenker argues that present tense is indexical, denoting the time of utterance. English present tense is a non-shiftable indexical, that is, it can only refer to the time of the actual speech act.⁷ In contrast, Russian present tense can be shifted. This means that it can be evaluated with respect to the context introduced by the attitude verb, as in (40) (Schlenker, 2003, p. 70, (56a)):

- (40) Petja_i skazal, čto on_i plačet.
 Petja_i said that he_i is-crying
 'Petja said that he was crying [at the time of his utterance]'

Here, the indexical present tense of *plačet* is shifted to denote the time of Petya's stating that he was crying. Schlenker (1999) gives an example of complements embedded under a non-attitude verb ((41, based on Schlenker's (5), p. 33), to demonstrate that the difference is semantic rather than syntactic:

- (41) Často slučalos', čto Miša plakal/*plačet.
 Often happened that Misha cried-IFPV/*is crying
 'It often happened that Misha cried/*is crying'

In these sentences, a complement clause with a past verb is embedded under a past verb which is not indexical. Just as Schlenker predicts, present tense cannot be used in the complement clause to denote simultaneity.

As for past tense in English and other SOT languages, Schlenker (2004) argues that some rule of morphological agreement is needed to account for (at least some) SOT cases.

⁷It should be noted that the present tense in English (as well as in Russian) can be shifted when embedded under future:

- (i) John will say that he is ill.

Thus, a separate rule has to govern future embeddings. One option, mentioned in section 3.1.4, is to treat future auxiliaries as modals.

Schlenker's theory is attractive for two reasons. First, it succeeds in explaining the well-known contrast between the behaviour of the English present tense in complement clauses (where it triggers a double-access interpretation) as opposed to other types of embedded clauses in a simple and elegant way. Second, it subsumes Russian present tense under a general class of shiftable indexical expressions, so Russian present tense does not stand out as an exotic phenomenon. In fact, Schlenker's theory implies that shiftable indexicals are the unmarked option in the system: their interpretation can depend on any context introduced in the utterance; by contrast, the interpretation of unshiftable indexicals is restricted by further limitations. In this sense, the behaviour of the Russian present can be viewed as the default case.

Babyonyshev & Matushansky (2006) (BM) propose an analysis of Russian present and past tense in SOT environments which follows Schlenker (2003, 2004) in claiming that Russian present tense is a shiftable indexical which can be evaluated either with respect to the context of utterance or with respect to some other context (yielding a simultaneous reading in a present-under-past embedding). However, they argue that some syntactic conditions are still necessary to describe the behaviour of the Russian simultaneous tense. As has been mentioned, Schlenker's theory involves an essentially semantic mechanism: the verb must be used in an attitude environment for the indexical tense to be able to shift. BM suggest that if an attitude context is sufficient for shifting Russian present tense "into the past", then frame adverbials denoting attitudes which were expressed in the past must produce the same effect. However, this is not the case. Consider the following contrasts (examples mine – O.K.):

- (42) a. Maša utverždala, što ona ljubit Ivana.
Masha claimed that she loves Ivan
'Masha claimed that she loved Ivan (at the time of claiming)'
- b. Mašino (prošlogodnee) utverždenie, što ona ljubit Ivana,
Masha's (last-year) claim that she loves Ivan
segodnja kažetsja absurdnym.
today seems absurd
'Masha's claim (made last year) that she loved Ivan [at the time of the claim] now seems absurd'
- c. Po (prošlogodnemu) utverždeniju Maši, ona ljubit
By (last-year) claim Masha-GEN she loves
Ivana.
Ivan.
'According to Masha's claim (made last year), she loves Ivan.'

In (42a) and (42b), present has a shifted interpretation. In (42a), present is interpreted as simultaneous to the matrix event, located in the past. Note

that in (42b), which involves embedding under an attitude NP (*utverždenie* 'claim'), present is interpreted as simultaneous to the event in the past described by the NP; importantly, the matrix T is present-tensed so it does not contribute to the shifting of the embedded present. If past is used in the NP complement, it will normally get a past-shifted interpretation:

- (43) Mašino (prošlogodnee) utverždenie, čto ona ljubila Ivana,
 Masha's (last-year) claim that she loved Ivan
 segodnja kažetsja absurdnym.
 today seems absurd
 'Masha's claim (made last year) that she used to love Ivan [before
 the time of the claim] now seems absurd'

However, in (42c), the present tense on the verb can only be interpreted as actual present, that is, Masha's feelings towards Ivan are located in the here and now, even if we assume that the claim was made in the past. Based on this observation, BM add a syntactic condition to Schlenker's treatment of tense: they argue that present tense must be c-commanded by the attitude predicate. Importantly, according to BM, this relation has to be realized in syntax proper, because at LF, attitude adverbials have the verb in their scope, just as attitude verbs or nouns do. Therefore, BM propose a mixed syntactic/semantic approach to Russian tenses. This is captured in "the shifting condition": "A shifted indexical is licensed by a c-commanding attitude report predicate". Both present and past tense can be indexical or shifted. When present tense embedded under an attitude verb is indexical, it produces a double-access reading, analogous to that found in English. When it is shifted, it denotes simultaneity. Likewise, an embedded past tense can be indexical and be evaluated as preceding the utterance time; it can in principle be simultaneous with a matrix past tense, but this possibility is limited by the properties of the matrix verb. When an embedded past tense is shifted, it denotes precedence with respect to the matrix event time.

However, note that BM's counterargument is based on the assumption that the attitude operator expressed by the adverbial c-commands the verb at LF. Thus, the analysis involves c-command relations at LF that are not present in surface syntax, something that is not allowed in many modern semantic theories, such as (Reinhart, 2006). If the attitude adverbials do not c-command the verb either on the surface syntax level or on the logical syntax level, the problem does not arise.

In a similar vein, Philippe Schlenker (p.c.) suggests that the adverbials in question are not analysed as attitude reports; rather, they are modalized expressions of the speaker's own thoughts (comparable to: "According to what I read, John will come tomorrow / possibly, John will come tomorrow", etc). I will return to this discussion in section 4.3.2.

Stechow (2003a,b); Kratzer (1998, 2006)

Stechow (2003a,b) and Kratzer (1998, 2006) present two similar accounts that emphasize parallels between tenses and pronouns and draw on some insights from work of Abusch and Schlenker. The starting point of (Kratzer, 1998) is an example by Irene Heim (class lectures) (her (4), p. 1):

(44) Only I got a question that I understood.

This sentence has a strict and a sloppy interpretation. In case of a strict reading, the speaker did not understand any of the questions that all the other people got, she only understood the one that she got. Under the sloppy reading, it is assumed that no one except for the speaker got questions they understood, but the speaker got a question understandable to her. Kratzer concludes that “the person and number features of the second *I* are not interpreted, and the pronoun has a bound variable interpretation” (Kratzer, 1998, p. 1).

However, there is a problem with the variable interpretation. 3rd person pronouns can be free variables: in that case, their index is left free, and they get their value from the context. In contrast, 1st and 2nd person pronouns cannot be free variables; they always have to be bound. Hence, the analysis cannot be the same as that of 3rd person pronouns. Kratzer’s solution is to introduce a new class of pronouns – “zero pronouns”, symbolized as \emptyset , that enter the derivation without any ϕ -features altogether, and get their pronunciation from an antecedent at PF. It follows that they must always have an antecedent. The LF of the sentence (44) corresponding to the sloppy reading can then be represented as (45):

(45) [Only I]₁ got a question that \emptyset ₁ understood.

Importantly, Kratzer detects locality constraints to the relation between the zero pronoun and its antecedent that do not apply for the third person pronouns as bound variables: the antecedent and the zero pronoun cannot be separated by more than one clause boundary.⁸ She compares that to the behaviour of the controlled PRO. Thus, there are syntactic conditions on the appearance of the zero pronoun.

Analogously, Kratzer argues that the inventory of English tenses consists of a present and a past tense (indexical tenses) and a zero tense. A zero tense is a lexically indexed variable that is bound by a local antecedent. In other words, a zero tense must always be anaphoric to the tense in the next higher clause. In Kratzer’s opinion, a zero tense is what we find in English complement clauses when a past-tensed verb gets a simultaneous reading. In these cases, the tense on the embedded verb only looks like past, but in fact receives its past morphology only at PF. It “has no presuppositions” – that is, it does not encode any temporal ordering.

⁸However, see (Cable, 2005) for a different view on Kratzer’s examples.

Consequently, it does not have an inherent past interpretation; its interpretation coincides with that of its antecedent (the matrix past).

The same line is pursued in (Kratzer, 2006), where Kratzer focuses on pronouns and analyses long-distance examples of “bound” 1st person pronouns, like (46):

- (46) We are the only people who are taking care of our children. (Kratzer, 2006, p. 10, (11))

Interestingly, Kratzer (2006) derives the sloppy readings of indexical pronouns in sentences like (44) and (46) by means of an agreement chain that links the two pronouns via intermediate nodes. The chain is divided into the following steps:

- Subject verb agreement
- Agreement between a predicative DP and its subject
- Agreement between a relative pronoun and its head
- Subject verb agreement
- Agreement between a verb and a possessive pronoun in the specifier position of the *verbaes* direct object. (Kratzer, 2006, p. 10)

As Kratzer points out, the only problematic links in this chain are the agreement links between a predicative DP and its subject (*we* and *people*), and between a relative pronoun and its head (*people* and *who*). The question is how 1st (or 2nd) person features can be transmitted via these links. Kratzer’s solution is based on the observation that according to many analyses, 3rd person features do not exist, and the corresponding forms are simply underspecified for person. In this case, NPs and DPs traditionally thought of as 3rd person could support 1st and 2nd person agreement.

A similar account of pronouns and tenses is presented in (Stechow, 2003a, 2003b). Stechow analyses the same type of interpretations for 1st and 2nd person pronouns as demonstrated in (44). However, while Kratzer argues that interpretable features cannot disappear on the way to LF, Stechow claims exactly that: in his model, features of semantically bound variables are deleted at LF, and so they are not interpreted. This is what happens to the feature *person* of the pronouns in examples like (44) in case of sloppy readings.

Stechow argues that the same mechanism of feature deletion is at the core of SOT effects. He claims that “Person, Mood and Tense are features on the verb that are checked by the features of the correspondent arguments of the verb, which are an individual, a world and a time variable” (p. 1). Attitude verbs, modal verbs and the future “will” are variable binders; they semantically bind the corresponding features on the embedded verb in their complements under agreement. Hence, when the feature Past is

semantically bound by the Past feature on the matrix attitude verb, it is deleted and not interpreted at LF.⁹

Stechow (2003b) argues against Kratzer's analysis, pointing out that it does not predict the behaviour of the Russian present tense, since it only deals with the English-type SOT where the features of the antecedent tense and the zero tense are identical. Stechow (2003a) attempts to solve this problem by introducing a "pres-Parameter" for Russian: "Russian verbs of attitude delete (LF) the feature *pres* at the temporal variable they bind, regardless what their tense (their temporal checkee) is" ((56), p. 18). However, in this case, the question arises why the mechanism of feature deletion is blocked for Russian past tense, which does not have a simultaneous reading of past analogous to the English one. The "pres-Parameter" only serves to ensure a simultaneous reading for the embedded present, but does not exclude the English-type agreement and deletion mechanism when the embedded tense is past. According to Stechow's model, past and present should be interchangeable in Russian complement clauses. On the other hand, if we assume that a different mechanism governs the behaviour of present in Russian, Kratzer's account seems to fare well for Russian: the only assumption needed is that Russian does not have a zero tense.

However, Kratzer's analysis also raises questions. A criticism of Kratzer's (1998, 2006) theory of 1st and 2nd person pronouns is found in (Khomitsevich & Reuland, 2007; Reuland, 2007). First, from the point of view of the contemporary syntactic theory, it is not optimal to have two sources for one and the same lexical item. Moreover, it goes against the Inclusiveness Condition (Chomsky, 1995) which views syntax as a combinatorial system with a strictly morphosyntactic vocabulary. The same argument can be applied to the inventory of temporal morphemes.

Moreover, many languages display underspecified anaphors, which do not inherit the feature makeup of their antecedents. Instead, they are spelled out without the relevant features. For example, in an analogue of (46) Russian would use the possessive anaphor *svoj* "self's", which is not specified for person, number or gender; it stays the same no matter what the features of the binder are. To account for this, Kratzer's model would have to be equipped with a sort of Optimality Theory ranking: the zero pronouns are spelled out as underspecified elements if these are present in the system, or otherwise as the next best thing – pronouns specified for the relevant features, which would be a serious departure from Kratzer's original idea. (Reuland, 2007) notes that the feature deletion approach might be more successful here; however, it appears too unrestrictive, since Kratzer (1998) shows that the positions in which 1st and 2nd pronouns get

⁹Thus, Kratzer's treatment of SOT is in the spirit of those accounts which introduce an SOT rule spelling out a zero tense (or present tense) as past, such as Comrie (1986); Stechow's account, on the other hand, is closer to the spirit of Ogiwara's (1989 et seq.) theory, in which past tense in SOT configurations is deleted by the SOT rule

sloppy readings are limited by locality conditions.

(Reuland, 2001) demonstrates that in many languages, e.g. Dutch, 1st and 2nd person pronouns can be locally bound. As explained in (Reuland, 2007), this analysis can be extended to cover examples like (44) and (46). Moreover, in order to distinguish between sloppy readings of sentences like (46), which, as Kratzer argues, exhibit locality effects, and some other cases where sloppy readings are derived via a logophoric mechanism, Kratzer has to introduce a special interface strategy rule similar to Reinhart's Rule I, to prevent the semantic logophoric mechanism from obviating the ban on feature transmission. In Reuland's version, an additional interface rule becomes unnecessary.

In the next chapter, I will provide an account of SOT in languages like English which bears some similarities to Kratzer's approach. I will show how temporal dependencies can be encoded via feature agreement chains – a solution that Kratzer proposes for pronouns but does not extend to tenses. However, I will show that it is possible to implement such an account without introducing a new member to the inventory of tenses, simply by applying the agreement mechanism already introduced for other linguistic phenomena.

Kusumoto (1999)

In her 1999 dissertation, Kiyomi Kusumoto provides an account of embedded tenses that aims to cover the whole variety of tense systems: “SOT languages”, like English, and different types of “non-SOT languages” like Japanese on the one hand and Russian and Polish on the other hand. Kusumoto's account is, as she admits herself, a “hybrid” between Ogihara's, Stowell's and Abusch's proposals, though relying somewhat more on the first one. However, in contrast to Ogihara, she does not postulate an SOT rule to account for simultaneous past tenses. In her model, tense morphemes are time variables that serve as arguments for predicates, which have a time argument slot in addition to the other arguments. Tense morphemes by themselves are semantically vacuous. Kusumoto denotes them as *past* and *present*. The tense meaning – the ordering of events described by verbs – does not come from the tense morphemes but from phonologically null tense operators that stand in a certain relation with the tense morphemes. These operators are denoted as *PAST* and *PRESENT*. When a tense morpheme occurs without a tense operator, it does not carry any meaning of ordering. It can then be considered an instance of “vacuous tense”. However, such occurrence is restricted: vacuous past must be locally c-commanded by an operator *PAST* (“locally” here means that no other tense operator should intervene). What this means is that when there is no tense operator in the embedded clause, the *past* morpheme can be licensed if it is in the scope of a *PAST* operator in the matrix clause. When this happens, a simultaneous past-under-past interpretation is de-

rived.

Under the definition above, such an interpretation must be available both for complements and for relative clauses. This is indeed what Kusumoto assumes, along the lines of Ogihara (1989, 1996). The difference between relative and complement clauses is derived in a semantic way. Not only vacuous past, but also “true” past can occur in all positions, embedded clauses included. This means that a *PAST* operator can be present there. In Kusumoto’s system, embedding one tense under another does not change its interpretation as long as the corresponding tense operator is available, because every tense operator is evaluated with respect to an overtly realized time variable (denoted as t^*), which is present in every clause. This time variable is set as speech time by default. Therefore, in relative clauses there is no restriction for the embedded past to be evaluated as independent of the matrix past, and be interpreted with respect to the utterance time. Complement clauses are different not structurally but semantically, because they are complements of attitude verbs. A proposition cannot be embedded under such a verb, because this results in a type mismatch (Kusumoto adopts a system where attitude verbs take as arguments properties of time, type $\langle i, st \rangle$, while a proposition is of type $\langle s, t \rangle$). So the embedded time variable must be bound by a lambda operator. This gives the past-shifted interpretation, but not the “later-than-matrix” interpretation, which is available for relative clauses when the embedded time variable is not bound. In fact, the vacuous past tense is also bound by a lambda operator in a complement clause, for the same reason, only we do not feel the difference in that case.

As for past in languages like Japanese and Polish/Russian, Kusumoto (1999) suggests that, just as SOT languages, non-SOT languages have two past tenses – a semantically vacuous variable *past* and an operator *PAST*. She argues against the view that non-SOT languages lack a vacuous past. Instead, she proposes a system where the constraints on the appearance of the *past* variable are different: it can only be bound by a *PAST* operator in the same clause.

The intuition behind this analysis is somewhat similar to the approach that I will be taking. The major difference is that Kusumoto’s account is semantic. However, the same idea can be formulated in syntactic terms: the temporal morphology is located on the verb, but the locus of tense interpretation is elsewhere – namely, in the T node. This would roughly correspond to Kusumoto’s distinction between tense operators and “vacuous” tense morphemes. Then, if the embedded T cannot provide a temporal interpretation, a higher T has to step in the breach. The difference is of course that the semantic relation Kusumoto proposes does not have to be (syntactically) local. In my model, a binding-like process between the two T’s ensures the locality of the temporal licensing.

An interesting part of Kusumoto’s account is her treatment of tenses in Japanese relative and adjunct clauses. As noted in section 3.1.5, Japa-

nese relative and adjunct clauses display a dependent interpretation of the present tense, in contrast to Russian where tenses in relative and adjunct clauses behave like they do in English (i.e. are interpreted as independent). Kusumoto proposes that the observed contrast is due to the structural properties of relative and adjunct clauses in Japanese. She analyses “Japanese ‘present tensed’ relative clauses as tenseless clauses more or less parallel to the English participle construction and Polish and Russian ‘present tensed’ relative clauses as full-fledged relative clauses parallel to English relative clauses. It follows that the temporal interpretation of ‘present tensed’ relative clauses in Japanese may be dependent on operators that embed them.” (p. 193). As for adjunct clauses, Kusumoto (1999) adopts the hypothesis put forward by (Geis, 1970) that adjunct clauses are structurally analysed as relative clauses.

This is how the explanation works. First, Kusumoto presents evidence from earlier research (Murasugi, 1991) proposing that Japanese relative clauses lack a C projection (a primary reason for this hypothesis is that they do not exhibit island effects (Kuno, 1973)). Further, she argues that when these clauses are present tensed, they do not have a T projection, either. As many researchers have suggested, the Japanese present tense is not a tense at all, but rather a tenseless form that displays default morphology which does not carry any temporal semantic input. Thus, traditional grammarians analysed the present tense morphology in Japanese as “imperfective endings”, “non-past endings” or “basic endings”; a similar approach is also developed in (Ogihara, 1995).

Finally, in English there are other differences between relative clauses and simple adjectival modifiers: relative clauses can have an overt relative pronoun, and there are also positional restrictions: relative clauses always follow the nouns that they modify, while simple adjectival modifiers can precede them. These differences are absent in Japanese: there are no relative markers, and all modifiers have to precede the noun independent of their structure. Hence, nothing tells us whether the modifier in an example like (47) is a relative clause or not except for the tense morpheme. Since this morpheme also “does not count”, nothing prevents Kusumoto from claiming that relative clauses are semantically parallel to English participial constructions, like in the first translation of (47).

- (47) Mariko-wa naiteiru otokonoko-ni hanasikaketa
 M-top cry-pres boy-to talk-past
 ‘Mariko talked to the crying boy’
 ‘Mariko talked to the boy who was crying’ (Kusumoto, 1999, p. 208, (147c)).

This English sentence, repeated below as (48a), has a dependent reading for the embedded verb, as opposed to (48b), which only has an independent reading for the verb in the relative clause (the boy is crying at the moment of speech):

- (48) a. Mariko talked to the crying boy.
 b. Mariko talked to the boy who is crying.

Kusumoto proposes that English participle clauses like the one in (48a) are tenseless and their temporal interpretation is provided by the matrix clause. The same analysis is thus valid for Japanese relative clauses.

In Russian (and Polish) present tensed relative clauses the situation is different. Kusumoto first argues that present verb forms in these languages (in contrast to past forms) do not display any specific temporal marking, and the inflection that they carry can be taken to encode agreement only (cf. (Jakobson, 1932)). In this, they are similar to Japanese. However, Kusumoto shows that Russian relative clauses necessarily have a fully-fledged CP system, and are structurally different from participles.

Kusumoto's analysis seems to show convincingly that the behaviour of Japanese relative and adjunct clauses is explained by their structural properties. If Japanese relative clauses lack CP and TP (or perhaps their TP is deficient in some way), by any account they have to have a dependent tense meaning. In the subsequent discussion, I will assume that this analysis is correct, and that the Japanese relative and adjunct clauses do not constitute a counterexample to a theory of SOT that relies on phenomena in complement clauses (and/or under attitude verbs). However, a full-fledged analysis of Japanese is outside the scope of my work; my account will primarily deal with Russian as opposed to English and Dutch.

3.3 Summary

In this chapter, I have introduced the notion of Sequence of Tense (SOT) and reviewed several accounts of SOT phenomena in English and other languages. Below I summarize the observations and conclusions that have resulted from my literature overview.

1. The theory of (Enç, 1987) was partly successful in establishing a syntactic binding mechanism for tenses. However, first, the theoretical apparatus of the classical BT has proved untenable for independent reasons; second, Enç's theory has a limited empirical coverage; importantly, it is ill equipped to deal with the whole wealth of data from non-SOT languages.
2. Abusch (1997) shows convincingly that a simultaneous reading of past under past can be derived in two ways. The first is described by an "independent" tense theory; in this case, the embedded tense is connected to the embedding tense via the general mechanisms operating in discourse, and the embedded tense is interpreted *de re*. The second mechanism establishes a special dependency relation between the two tenses. In what follows, I will show that this intuition is on

the right track, and that it correctly explains the contrast between the simultaneous past in English and Russian: in Russian, only a *de re* simultaneous past is present, while in English, an additional mechanism seems to be at work. However, I will take a different stand with respect to what kind of mechanism this is.

3. While Abusch's independent tense theory can be made to work for the Russian past, Schlenker's (2003, 2004) approach to Russian present as a shiftable indexical covers the rest of the Russian data. I have also addressed the criticism of Schlenker presented by Babyonyshev & Matushansky and have shown that the theory can still stand up to it.
4. My account shares an important insight with (Kratzer, 2006), namely the idea of agreement chains transmitting long-distance relations via a number of local steps. Kratzer applies this approach to 1st and 2nd person pronouns, while I will apply it to tenses. However, in contrast to Kratzer, I will not postulate a special agreement mechanism dealing specifically with my data; rather, I will rely on the feature chain formation approach developed in (Pesetsky & Torrego, 2001 *et seq.*, Reuland, 2001, 2005) that already has been shown to work for other linguistic phenomena.
5. Finally, I have demonstrated that Kusumoto's (1999) account explains the (apparent) difference between the two "types" of non-SOT languages, represented by Russian and Japanese.

Chapter 4

The analysis of SOT in English and Russian

In this chapter, I develop my analysis of SOT in “SOT-languages”, like English and Dutch, as compared to non-SOT languages, like Russian. In section 4.1, I address the problem of the simultaneous past readings in SOT languages, in particular, the existence of seemingly vacuous past uses, and argue that the “truly vacuous” past should be treated separately from other uses of the past tense. In section 4.2, I present my own account of the SOT phenomena, which is based on an Agree relation between tense morphemes. Some remaining issues are addressed in section 4.3. Section 4.4 summarizes the contents of the chapter.

4.1 The nature of simultaneous past

In this section, I focus on the contrast between SOT and non-SOT languages; the examples for the former group that I will be using come mostly from English and Dutch. The main concern of this section will be the nature of the simultaneous past in English (and other SOT languages) as opposed to Russian. I will follow the account presented in a more sketchy fashion in (Khomitsevich and Reuland, 2007). There, it was proposed to treat the “truly vacuous” past as instances of an irrealis use of the past morpheme. Further, it was argued that although cases of simultaneous past are found in Russian, their interpretation is different from English. Finally, it was proposed that simultaneous past readings in English are derived via a syntactic binding mechanism between the tenses in the embedding and embedded clause. In what follows, I will elaborate on this idea, presenting more arguments in favour of this view, as well as a concrete implementation of it in contemporary syntactic terms.

4.1.1 Simultaneous past in SOT languages

The “truly vacuous” past

In what follows I will address the nature of simultaneous past in SOT languages like English. One of its most striking properties is the existence of the “truly vacuous” past tense, as in the “last meal” examples, which were presented in section 3.1.2; they are repeated below as (1):

- (1) a. John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together (Abusch, 1988, p. 2, (6)), originally from (Kamp & Rohrer, 1984).
- b. We decided to tell the prosecutor tomorrow that we were talking to him reluctantly (Enç, 2004, p. 205, (6)).

In these sentences, the deepest embedded past (*were*) does not refer to a past time at all – it precedes neither the moment of speech nor any other evaluation point introduced in the sentence; rather, it refers to a time which follows the moment of speech. Such sentences have been viewed as a counterexample to (Enç, 1987, 2004), whose theory cannot explain the appearance of past in the most embedded clauses like (1), since in her model, past morphemes are always anchored as past with respect to the time of utterance. Abusch’s “independent” tense theory is also unable to deal with them. Instead, many other authors have proposed that the simultaneous past tense in SOT configurations is in fact a null (vacuous) tense that only looks like past. (Thus, in Ogihara’s theory, past tense on the verb is deleted at LF; in Abusch’s feature transmission model, the “pastness” of the embedded verb also “does not count”, etc).

However, these examples still pose problems. Ogihara (1996) cites examples like (12a) from section 3.1.4, repeated here as (2) (similar examples are found in (Abusch, 1997), given here in (3) and (4)). Arguing against Abusch’s (1988) theory of feature transmission, he claims that since in Abusch’s theory tense feature transmission is only possible from an intensional predicate, tense in the deepest embedded clause must depend for its interpretation on the verb *said*. But this is an undesirable result, because the time of the fish’s being alive is not ordered with respect to the time of saying: the sentence can be understood as meaning that John was going to buy a fish that would be still alive at the time of his buying it. Similarly, according to (Abusch, 1997), examples (3) and (4) can mean that Sue did not have any particular man in mind yet but rather expressed an opinion that her future husband would love her.

- (2) John said he would buy a fish which was still alive.
- (3) Sue believed that she would marry a man who loved her. (Abusch (1997), (30), p. 17)

- (4) Sue expected to marry a man who loved her. (Abusch (1997), (31), p. 17)

Kusumoto (1999) suggests that the problem with examples like (2) can be solved if the auxiliary *will/would* is treated as an intensional operator, on a par with verbs like *say* (this idea comes from a later version of Abusch's account in (Abusch, 1997)). Indeed, many analyses treat the future auxiliary as a modal, which means it can be classified as an intensional predicate. However, there is no intensional predicate embedded under *expected* in (4), unless we assume that a covert intensional operator is present whenever a semantic shift to the future occurs. (Note that in any case, Ogihara's objection seems valid for theories that rely on attitude contexts (and not on intensional contexts in general), such as Stechow's).

More problematic examples can be added to this discussion. Consider the following sentences, which involve embeddings in unreal conditionals:

- (5) a. If a mugger attacked me, I would tell him that I did not have any money. (Khomitsevich & Reuland, 2007, (23))
 b. If you were king, you could cut off the heads of everyone who *offended* you. (Partee, 1973, p. 608, (26))

In these sentences, the matrix verb refers to some hypothetical moment in the present or future; no verb in the sentence makes a reference to any past moment. Still, the verb in the complement clause in (5a) and in the relative clause in (5b) carries past morphology. (5a) bears a resemblance to the "last meal" examples, since the "vacuous" past verb is embedded under *would*. However, here there is no higher tense which would refer to a past time and "transmit" the past tense down to the most embedded verb. Formally, it is possible to argue that *would* carries a past feature (*would* being a past tense of a modal *woll*, with the present form *will*, as Abusch (1997) suggests. But in this case, it would have to be an instance of vacuous past itself, since it does not refer to a past moment, and then it is not clear what licenses this past feature. Semantically, *would* in counterfactuals is an irrealis form and not a past form. But then, nothing conceptual forces the verb embedded under it to be past.

The solution for this puzzle that I propose is to separate cases when past does not encode temporal precedence from those when it does, while still maintaining that they have a common denominator on a more abstract level. If we treat the "last meal" sentences and similar cases separately, the problem is then to establish why the embedded verbs in them carry past morphology. Note that there are two environments where "vacuous" past occurs: future-shifted embeddings and clauses embedded under counterfactuals like conditional clauses. What these contexts have in common is an irrealis semantics. It is clear that conditionals are counterfactual; as for the future, it has been long argued that it should be viewed as modal rather

than temporal (Enç, 1996, a.m.o). Thus, both future and conditionals provide an irrealis context. I argue that what we have there in SOT languages such as English or Dutch is an irrealis use of the past morpheme.

Past and irrealis

In this section, I will discuss the relation between past and irrealis. In English, the form carrying past morphology (the ‘-ed’ form for regular verbs) is ambiguous between a past tense and irrealis mood, or past subjunctive, as it has been termed in the traditional grammar. Past subjunctive is used in the if-clauses of conditional constructions and in other contexts like “I wish...”, “it’s time...”, “if only...”; in most cases, it is identical with the simple past indicative forms:

- (6) a. If it rained, we wouldn’t go out.
- b. I wish it didn’t rain!
- c. If only it didn’t rain!
- d. It’s (high) time I went to work.

The only remnant of a special past subjunctive paradigm is the form *were*, which can be used for all persons and numbers in the subjunctive, but not in the past indicative:

- (7) a. If John was/were rich, he would buy himself a Ferrari.
- b. I wish I was/were rich!
- c. John was/*were rich.

Abusch (1997, p. 19-20, fn. 13) proposes to use this contrast in order to show that the deepest embedded past in examples like (1) is really past and not subjunctive: in such contexts, *were* cannot be used with 1st and 3rd person singular. However, there is no straightforward correlation between the distribution of subjunctive and *were*. As noted in (James, 1986), there are certain constructions where only *was* and not *were* can be used. For instance, this is what happens in a construction like (8a), which is synonymous to (8b). Although there is no difference in meaning, the first sentence only allows the form *was*, and the second – both *was* and *were*. The construction “it’s time...”, illustrated in (8c), allowed both forms in earlier English, but not in the modern usage.

- (8) a. He’s behaving like he was/*were sick.
- b. He’s behaving as if he was/were sick.
- c. It’s time he was/*were leaving. (James, 1986, p. 108)

On the basis of this evidence, James (1986) argues that in modern English the differences between moods are no longer recognized, and “to the modern intuition, *were* is merely an alternant of *was* which is restricted

to particular environments, these including the word *if* but not the word *like*" (p. 108–109).

Similarly, (Comrie, 1985) points out that the distinction between the past forms of *be* in past contexts or counterfactuals only holds for some speakers. He also cites a context where there is no such distinction, namely polite requests like *I just wanted to ask you if you could lend me a pound* (Comrie, 1985, p. 19). Here the past verb *wanted* refers to the present moment and not to the past. The function of the past here is to indicate politeness, which can be also done by using an form like *I would like* instead of *wanted*. However, it is impossible to use *were* with 1st/3rd singular in such a construction. I conclude that the use of English past in the irrealis function cannot be restricted to the environments where the *was/were* alternation is possible.

In Dutch, which behaves like English with respect to SOT, the ambiguity between past and irrealis is even more striking. The morphological past form and the *zou*-form (the analogue of the English *would*) are interchangeable in conditionals (in contrast to English):

- (9) Als het niet zou regenen / regende, zouden we naar buiten
 if it not would rain / rained, would we to outside
 gaan / gingen we naar buiten.
 go / went we to outside
 'If it did not rain, we would go out.'

Moreover, Dutch displays cases where past can appear in a root clause with an irrealis meaning. These are examples of the so-called *imparfait préludique* – contexts of children's games when a player suggests the conditions of the game:

- (10) En dan was ik de vader en jij de moeder.
 and then was I the father and you the mother
 'Let's pretend I am the father and you are the mother'

Note that the conjunction *dan* is not normally used with past; it can only be used in relation to present events. The phenomenon illustrated in (10) is not specific to Dutch; the same use of the imperfect, along with other modal uses, is reported for Romance languages in (Giorgi & Pianesi, 1997).

Other studies have shown that the connection between past and irrealis is not accidental: there is an intrinsic affinity between the two. Thus, (Steele, 1975) shows that in various typologically unrelated languages (Garo, Sino-Tibetan; Chipewyan, Athapaskan; Old Marathi, Indo-Aryan, etc), the same morpheme is used to mark both past tense and irrealis. On the basis of this observation, she suggests a reconstruction for Proto-Uto-Aztecan where the past and the irrealis morphemes are descended from the same element (**/ta/*) rather than two homonymous elements. She notes that also in English, the morpheme *-ed* can indicate

either past tense or irrealis, and claims that past and irrealis have in common a semantic primitive, which she calls DISSOCIATIVE. This primitive refers to the marked status of past and irrealis; both represent the opposition to the unmarked option in the system: past is opposed to present, and irrealis – to realis. Steele (1975) argues that this semantic primitive is a universal.

(Iatridou, 2000) notes that in Modern Greek, as well as in English, counterfactuals (conditionals, irrealis wishes, etc) display past morphology which does not receive a past interpretation. She terms these uses “fake past” but goes on to argue that this use of past morphology is not a case of “accidental homophony”. Apparently, “fake past” uses are too widespread crosslinguistically to be a mere coincidence. Iatridou (2000) claims that the morpheme in question has a constant meaning, but this meaning can be realized in different domains. She defines this meaning as the “exclusion feature” (ExclF) – a feature that serves to exclude the current coordinates of the speaker. This feature can range over times or over worlds. When it ranges over time, it generates the meaning of past because past is “exclusion of” present (assuming that future is not an independent tense, but rather is included into the present domain or represented by a modal form); when it ranges over worlds, it gives irrealis as exclusion of the actual world.

Following Iatridou (2000), I propose that the (morphological) past form in English and Dutch has two distinct uses, as past tense and as irrealis (subjunctive) mood, which can ultimately be reduced to one meaning component. Thus, one form, with one underlying interpretation (which I take to be Iatridou’s “exclusion” interpretation) – we can call it the ED-form to avoid confusion – serves two purposes: it can act as a tense, expressing temporal precedence (pastness), and as a mood.¹ How exactly this form is interpreted depends on the particular context that it occurs in. I argue that the “truly vacuous” past form found in the “last meal”-examples in English is in fact an instance of the irrealis use of the ED-form, and not of its past use.

The question of course arises what is the distribution of this irrealis form and why. Abusch (1997) notes that a simple modal is not a valid context to licence the subjunctive. She gives an example in (11a), where *were* is ungrammatical; using *was* here is allowed but does not have a simultaneous reading (it has to refer to the past); the same holds for (11b),

¹Note that English is traditionally supposed to also have present subjunctive forms, which largely coincide with present indicative forms but can be distinguished from them in some contexts, such as those in (i)-(ii):

- (i) They prefer that she come (James, 1986, p. 2).
- (ii) They insist that it be so (James, 1986, p. 28).

I leave an analysis of these forms and of the consequences that it might have for our understanding of the present tense for future research.

where the modal is in a matrix clause. However, in (11c) the simultaneous ED-form is fine.

- (11) a. * If anyone calls, I will/might say that I were busy. (Abusch, 1997, p. 19, fn. 13)
 b. What will you say if someone calls? – I will/might say that I am/*was busy.
 c. If anyone called, I would/could say that I was busy.

Thus, a so-called “unreal conditional” (as in (11c)) is a type of context that can licence the irrealis ED-form, while a “real” conditional or a modal cannot. I propose that “future-in-the-past” also belongs to the first type of contexts.

SOT languages: Towards a proposal

I have offered to distinguish cases of embedded past which do not denote precedence to any evaluation point from other instances of embedded past. In effect, now we are left with the cases where embedded past refers to an event(uality) preceding the utterance time. Thus, a theory of interpretation of embedded tense needs not postulate a vacuous embedded past; we can simply assume that the simultaneous past tense is interpreted with respect to utterance time. A theory that develops this assumption in the most straightforward manner is Abusch’s (1997) “independent” (extensional) theory, which was summarized in section 3.2.2. In short, this theory takes tenses to be referential expressions which are evaluated with respect to utterance time, but can be anaphoric to other tenses; it is then claimed that past tense embedded under attitude verbs is interpreted *de re*. This means that the embedded past tense denotes some past time in the real world, but in the belief of the matrix subject it is substituted by another time to which she bears an acquaintance relation. The absence of a forward-shifted reading is accounted for by the introduction of an Upper Limit Constraint, which makes the attitude time an upper limit for the interpretation of tenses in the attitude context.

I argue that this model of tense interpretation can (and should) be sustained; however, it is still not sufficient to explain the behaviour of embedded tenses in English. My reason for thinking so comes from comparing data from English and Russian. In the next section, I will show that simultaneous past is available in Russian, however it does not carry the same meaning as the English embedded past can have. On the basis of these data, I will argue that the simultaneous past tense in Russian is a *de re* tense which is anaphoric to the matrix tense, in line with Abusch’s independent theory. However, the simultaneous past tense in English is obtained via a binding process between the embedded and the matrix tense morphemes. Note that the *de re* option is still available for English, but the bound option does not exist in Russian.

These two options for tense interpretation present a nice parallel with the interpretation of pronouns, which can be bound or assigned the same value as a result of accidental covaluation (see section 2.2). The covaluation mechanism operates on a discourse level, while binding is a syntactic process. Similarly, the “independent” tense theory provides a general mechanism for linking tenses in the discourse. On the other hand, the binding relation that I am going to propose is a syntactic mechanism which is specific to the constructions that I am discussing.

4.1.2 Simultaneous past in Russian

The data

Russian is often called a “non-SOT” language, and described as having no simultaneous past. If this were true, it would present a perfect opposition to English and other SOT languages, which would be easy to explain by means of an SOT rule manifesting itself in English but not in Russian. However, this is not the case: simultaneous past in Russian is not as ubiquitous as it is in English, but it is not impossible. In section 3.1.5 I cited an example of past receiving a simultaneous interpretation in a complement of a perception verb from (Altshuler, 2004), repeated here as (12).

- (12) Dina videla, čto/kak voda lilas’/l’jôtsja iz vedra.
 Dina saw that/how water poured/pours from basket
 ‘Dina saw that/how the water was pouring from the basket’

Altshuler (2004) does not provide any theoretical explanation for the data. In fact, cases such as these have been discussed in literature for some time. The existence of simultaneous past in Russian was noticed in (Boeck, 1957, 1958) and (Costello, 1961). In (Boeck, 1957) it is claimed that in complement clauses with the conjunction *čto* “that” past tense with the meaning of simultaneity is not excluded but it is not used very often. However, with the conjunction *kak* ‘how’ the past and present tenses are interchangeable.

(Boeck, 1958) suggests that only past can be used with *happen*-type verbs, while only present is normally used in case of indirect speech (in the narrow sense of the word, i.e. after verbs of speech, especially with the conjunction *li* “whether”). However, both tenses are used with matrix verbs of perception, emotion and thought (*Wahrnehmung, Gemütsbewegung und Denktätigkeit*). Past is freely interchangeable with present only in the case of the conjunction *kak*. With other conjunctions (*čto* and its near-synonyms *budto, točno*) and in case of indirect questions present is preferred.

The same conclusions are reached by (Costello, 1961), who provides more examples of the use of simultaneous past. He also notes that:

- present is used more often than past with verbs of speech and mental activity;

- present and past are interchangeable with verbs of sense perception;
- the use of past stresses “the ‘real’ character of the fact stated in the dependent clause” (p. 494). He adds that the present is “more vivid” than past, and that it “involves an element of subjectivity” (p. 495).

All these facts are analysed in (Barentsen, 1996), a paper which offers a comprehensive empirical study of the use of tenses in perception context in Russian. Barentsen provides more examples of the use of tenses in complements of perception verbs. The fact that present and past are interchangeable in this environment is illustrated by the following example (I do not represent the aspect marking in Barentsen’s glosses):

- (13) Ona gľadela, kak size kol’ca ot sigary Azazello uplyvali
 she looked how grey rings from cigar of-Azazello flew-away
 v kamin i kak kot lovit ix na konec špagi.
 into fireplace and how cat catches them on end of-sword
 ‘She watched as Azazello blew smoke-rings at the fireplace and the
 cat spiked them on the end of his sword’ (Barentsen, 1996, p. 21,
 (18), from the novel *Master i Margarita*)

Barentsen suggests that there are two opposite groups of verbs that take embedded clauses: verbs with the meaning “to happen”, where only past tense can denote simultaneity, and speech verbs (which introduce indirect speech in the restricted sense of this term), where present is obligatory. Halfway between these groups are perception verbs, for which present and past are interchangeable.

Barentsen proposes that across these verb groups, “there is a decrease in the presupposed ‘reality’” of the situation described by the embedded clause. In case of *happen*-type verbs and to some extent perception verbs, the situation is “presented as part of objective reality”, and in this case, past is needed. In contrast, when the matrix verbs are of the type “think”, “say”, etc, the situation is associated with the “‘act of mind’ of the speaker, observer etc”, and in this case present tense is used (pp. 21-22). He also introduces the notion of “retrospective” vs. “narrative” use of the past tense. In the first case, the situation in the embedded clause is presented “from the point of view of the present moment” (the moment of utterance) (p. 36). In the second case, the relations between past events are more important than their relation to the present moment.

Another observation is the availability of two conjunctions used after perception verbs in Russian: *čto* ‘that’ and *kak* ‘how’. Only *kak* is used with verbs of “active” perception (*smotret’* ‘look’, *slušat’* ‘listen’), with others the two conjunctions are interchangeable. Barentsen defines the difference in the meaning of these conjunctions by comparing them to the English that-clauses and “sensory verb complements” (Kirsner & Thompson, 1976). The use of *kak* implies that the embedded clause situation is “not a state and

that it is maximally included in the scope of the act of perception”; “the action of the dependent clause is thought of as going on during the perception”. By contrast, “*čto* indicates that what is actually perceived functions as a clue giving rise to the thought of the situation referred to”. That is, the relation between the act of perception and the situation described in the embedded clause is “indirect”.

Quantitative data, which Barentsen takes from two corpora of modern Russian, show that while *čto*+present tense is a predominant use to express the meaning of simultaneity, when *kak* is used, past tense occurs quite often. Barentsen concludes that in *kak*-clauses both tenses are interchangeable.

Barentsen proposes that the reason for this behaviour of tenses lies in the properties of perception verbs. He notes that the semantic of these verbs presupposes simultaneity, and does not usually imply precedence or posteriority. For instance, in contrast to all other embeddings, past perfective forms in complements of perception verbs can be used with the meaning of simultaneity, *not* precedence:

- (14) On počuvstvoval, kak u nego poxolodel zatylok.
 he felt how with him turned-cold back-of-head
 ‘He felt a chill run up his spine’. (Barentsen, 1996, p. 25, (22))

Note that this is not the only interpretation of perfective under perception verbs. Compare (14) with (15):

- (15) Prixožu. Okno otkryto. Ja srazu počuvstvoval, čto
 I-come-in window open I at-once felt that
 čto-to slučilos’
 something happened
 ‘I went in. The window was open. I felt at once something had happened’. (Barentsen, 1996, p. 25, (24))

Here the meaning is one of precedence. This distinction seems to be governed by the choice of conjunction: under the conjunction *kak*, perfective will always mean simultaneity. Under *čto*, it will usually mean precedence, although this conjunction is more versatile. Compare:

- (16) Na rassvete ja počuvstvoval, čto po spine moej prošel oznob.
 At dawn I felt that along back mine passed shiver
 ‘At dawn I felt a fit of shiver run across my back’. (Barentsen, 1996, p. 28, (29))

In fact, specific properties of perception verbs have been noticed for other languages as well. Kusumoto (1999) reports the following example, due to Barbara Partee (p.c.):

- (17) Elliott observed/noticed/perceived that Josephine got hurt. (Kusumoto, 1999, p. 101, (12))

In (17), the embedded past can be understood to have a simultaneous reading, even though the embedded predicate is eventive and not stative. Barbara Partee (p.c. to Kiyomi Kusumoto) also notes that some times genuine belief contexts can have a similar interpretation (Kusumoto, 1999, p. 102, (16)):

- (18) a. I thought the glass fell by itself. I didn't know you pushed it.
 b. He didn't realize that his car hit the curb.
 c. The pilot was sure that the plane landed in the correct spot.

Partee concludes that these examples express belief based on perception and not simply opinions.

What is special about perception verbs is that the perception of an event is normally simultaneous with the event itself; we can see or hear or feel something only while it is still happening or is present in front of us, afterwards we can only perceive the consequences of what has happened. This is why matrix perception verbs strongly favour a simultaneous reading of its complement. Importantly, the grammar allows this simultaneous reading to be realized when the embedded verb has past tense.

These data suggest that simultaneous past-under-past constructions are more widespread in Russian than is usually claimed. To check this, I searched the online National Corpus of Russian language (Nacional'nyj korpus ruskogo jazyka, available at <http://www.ruscorpora.ru>) for the relevant data. The search yielded many examples where past tense embedded under past can easily be understood as simultaneous; moreover, many of the examples were disambiguated by the context to favour the simultaneous reading. This is true of sentences (19)-(22):

- (19) Vse my znali, čto ona žila v obščezitii teatra i čto
 All we knew that she lived in dorm of-theatre and that
 ej negde bylo prinimat' gostej.
 to-her nowhere was to-accept guests
 'We all knew that she was living/had been living in the theatre's dorm and she had/had had nowhere to invite guests' (simultaneous or past-shifted reading)
- (20) Čerez preemstvennost' naučnyx teorij nauka i
 through succession of-scientific theories science actually
 dvižetsja po puti ot neznanija k znaniju – nedarom
 moves along way from not-knowing to knowing – not-in-vain
 Njuton govoril, čto stojal na plečax gigantov.
 Newton said that stood on shoulders of-giants

'It is through succession of scientific theories that science moves along the way from not-knowing to knowing: that is why Newton said that he was standing/??had been standing on the shoulders of giants'. (simultaneous reading)

- (21) Ona bespokojno šarila vokrug sebja rukami, i on
she anxiously fumbled around self with-hands, and he
dogadalsja, što ona iskala ubežavšuju Murku, v kotoroj
guessed that she searched-for run-away Murka in whom
nuždalas' vsjakij raz, kogda popadala v zatrudnitel'noe
needed every time when got in difficult
položenie.
situation

'She was fumbling around anxiously with her hands, and he guessed that she was searching/*had been searching for the run-away Murka [a cat's name], who she needed every time when she was in a difficult situation'. (simultaneous reading)

- (22) Sčitalos', što ej šli kostjummy, kotorye ona nosila
was-considered that to-her fitted suits which she wore
so studenčeskix let.
since student years

'It was considered that suits, which she had been wearing since the time when she was a student, fitted/??had fitted her well'. (simultaneous reading)

In all these sentences, past tense in the complement clause can be replaced by present tense with no change in meaning. However, past tense sounds perfectly natural in these contexts.

To further test the availability of simultaneous past readings for complement clauses in Russian, I conducted a speaker questionnaire, asking speakers to judge sentences with a past verb in the matrix clause and a past complement clause introduced by the conjunction *što* 'that'. I used two groups of verbs in the matrix clause – verbs of thinking and factive verbs. The speakers were asked to judge whether the activity described by the embedded verb should be interpreted as preceding the matrix activity, simultaneous to it, or whether both are possible. It turned out that most speakers allowed simultaneous interpretation for all the sentences, including those with non-factive verbs (verbs of thinking) in the matrix clause. On the other hand, sentences with complements where present tense was embedded under past were judged unanimously as simultaneous, and those with a past perfective verb – as denoting precedence. To sum up, while present tense in a complement clause embedded under matrix past unambiguously triggers a simultaneous reading, there is no such dependency between embedded past and a past-shifted reading (or a simultaneous reading).

The interpretation

In the previous section, I have established that simultaneous past does indeed exist in Russian, although it has been claimed to be a non-SOT language. The question now arises what implication this fact has for the non-SOT status of Russian. In this section, I will show that the simultaneous past found in Russian is not analogous to the simultaneous past that can be found in English “SOT configurations” like (23), and which, as I will show below, is the consequence of a “binding” relation between tense morphemes.

(23) John said that Mary was pregnant.

To begin with, let me show that Russian does not have an analogue of the English “last meal-sentences” (sentences in which past tense is embedded under an infinitive or *would*, which is in turn embedded under past). In these examples, an instance of past tense does not have to precede the utterance time, it can even follow it because it is only ordered as simultaneous with respect to its closest higher tense. The relevant examples (the first two are from section 3.1.2, the last one from section 4.1.1) are given below:

- (24) a. John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together. (Abusch, 1988, p. 2, (6), originally from Kamp & Rohrer, 1984)
- b. We decided to tell the prosecutor tomorrow that we were talking to him reluctantly. (Enç, 2004, p. 205, (6))
- c. John promised to say that he did not know anything about the crime. (Ogihara, 1989, p. 96, (104))
- d. Sue expected to marry a man who loved her. (Abusch (1997), (31), p. 17)

In the previous section I proposed that the deepest embedded past in such cases is an instance of irrealis use for the past (ED) morpheme. That such a use is impossible in Russian can be demonstrated if we try to construct Russian sentences analogous to the English sentences in (24). Intuitively, simultaneous past in Russian is most easily available when embedded under a perception verb, so (25a) is grammatical under a simultaneous reading. However, simultaneous past is out in (25b).

- (25) a. Maša uvidela, kak Ivan gotovit/gotovil obed.
Masha saw how Ivan cooks/cooked dinner
‘Masha saw Ivan cooking dinner’.

- b. My dumali, čto zavtra Maša uvidit, kak Ivan
 We thought that tomorrow Masha will-see how Ivan
 gotovit/*gotovil obed.
 cooks/*cooked dinner.
 ‘We thought that tomorrow Masha would see Ivan cooking dinner’.

However, one might argue that this is because Russian does not have future in the past (a form analogous to the English *would*). Russian uses simple future instead. Thus, no past features are inherited by the embedded future. To avoid this problem, we can use an infinitival complement. Although examples with perception or factive verbs are not so easy to construct, the sentences in (26) may be the case at hand:

- (26) a. My poslušali, kak Lena delaet/delala doklad.
 We listened how Lena makes/made talk
 ‘We listened to Lena giving a talk’.
- b. My rešili zavtra poslušat’, kak Lena delaet/*delala doklad.
 We decided tomorrow to-listen how Lena makes/*made talk

‘We decided to listen to Lena giving a talk tomorrow’.

(26a) is a construction with a perception verb and the conjunction *kak*. This is the type of construction where past tense is widely believed to be used on a par with the present tense to render a simultaneous reading. However, in (26b), where the event denoted by the embedded infinitive is located in the future, past tense is emphatically out. It is worth noting that if we remove the adverb “tomorrow” and allow for the deepest embedded verb to refer to the past as well, then past is possible, although not quite as natural as present.

- (27) V tot den’ my rešili poslušat’, kak Lena delaet/?delala doklad.
 In that day we decided to-listen how Lena makes/?made talk

‘On that day we decided to listen to Lena giving a talk’.

Thus, past tense can be simultaneous with an infinitival complement embedded under a past verb, but only if all the relevant eventualities are located in the past.

As for sentences like (24d), we can use neither past nor present indicative to yield the intended meaning of its Russian counterpart:

- (28) Maša xotela vyjti замуž za človeka, kotoryj ljubil/ljubit ee.
 Masha wanted to-get married to person which loved/loves her
 ‘Masha wanted to marry a person who loved her [at that time] /
 loves her [at the time of speech]’

With indicative, the NP *človeka* can only mean a particular (specific) person; so, it can only be interpreted *de re*. For the intended *de dicto* reading, we have to use the subjunctive:

- (29) *Maša xotela vyjti zamuž za človeka, kotoryj ljubil by ee.*
 Masha wanted to-get married to person which love SUBJ her
 ‘Masha wanted to marry a person who loved her [a prospective person]’

This provides indirect evidence for the fact that an irrealis/subjunctive form is used in the relative clause in the English (24d), as well as other analogous examples.

Now I turn to the existing cases of simultaneous past in Russian. Let us recollect the contexts where the simultaneous past is more likely to be used in Russian. From the overview of the data, the following hierarchy arises: if the matrix verb is a perception verb, present and simultaneous past are virtually interchangeable in complement clauses. If the matrix verb is a factive verb like *znat’* “know” or *dogadat’sja* “realize”, simultaneous past is also widely used. With other types of matrix verbs (for example, *dumat’* “think” and *govorit’* “say”) simultaneous past is occasionally possible. According to intuitions in (Boeck, 1957), it is least possible in reported speech (under verbs of saying).

Intuitively, the suggestion in (Barentsen, 1996) that simultaneous past presents the event “as part of objective reality” seems to be on the right track. This follows from the semantics of sense perception verbs and factive verbs, since in sentences with these verbs the speaker commits herself to the truth of the complement. In contrast, verbs of speech or belief shift the focus to presenting the state of affairs as existing in the matrix subject’s mind. Cases of simultaneous past in Russian are more common with the former verb types; it seems that the simultaneous past helps to present the situation from the actual speaker’s rather than the matrix subject’s perspective. For temporal interpretations, this means that the embedded past tense is evaluated with respect to the time of the actual speech situation, not the time of the matrix eventuality. This picture is compatible with Abusch’s (1997) “independent” tense theory: as discussed in section 3.2.2, Abusch (1997) claims that tenses can be interpreted *de re*, in which case an embedded tense is dependent on the actual utterance time for its interpretation.

I propose that this is the only interpretation available for simultaneous past in Russian.² In what follows, I will go over several examples that support my conclusion. I will show that an eventuality denoted by simultaneous past in Russian is always perceived by the speaker as preceding the utterance time, that is, it is interpreted as “real” past. In contrast,

²The same conclusion is reached independently by (Altshuler, 2005), who also applies Abusch’s (1997) theory to Russian.

in English simultaneous past can be evaluated with respect to the matrix time, without relating it to the utterance time.

First, there is a difference between English and Russian in reporting past statements with individual-level predicates and generic statements. It is well-known that such statements are subject to the so-called “lifetime effect” (see e.g. Babyonyshev & Matushansky, 2006 and the references cited there), illustrated here by (30).

- (30) a. Gregory had blue eyes. (Musan, 1997)
 b. U Ivana byli golubye glaza.
 at Ivan were blue eyes
 ‘Ivan had blue eyes’

These sentences imply that Gregory and Ivan are both dead (or otherwise, they have somehow changed their eye colour). A related phenomenon are generally accepted facts like physical laws and laws of nature, in other words, statements that have “universal temporal validity” (Comrie, 1986), which cannot be made in the past tense:

- (31) a. The Earth is round/flat.
 b. # The Earth was round/flat.
 c. Zemlja vraščaetsja vokrug Solnca/Luny.
 Earth rotates around Sun/Moon
 ‘The Earth goes around the Sun/Moon’
 d. # Zemlja vraščalas’ vokrug Solnca/Luny.
 Earth rotated around Sun/Moon
 ‘The Earth went around the Sun/Moon’

However, if such sentences are embedded under verbs of attitude, then in SOT languages like English or Dutch, both past and present tense can be used (though for some speakers present tense is preferable in case of true statements, and past is preferable if the belief expressed is not true):

- (32) a. When he was five years old, John thought/said/knew that the Earth is/was round.
 b. When he was five years old, John thought/said/imagined that the Earth was/is flat. (examples based on Comrie, 1986, p. 285, (73)-(75))
 c. Toen Jan vijf was, wist hij al dat de aarde rond
 when John five was, knew he already that the earth round
 is/was.
 is/was
 d. Toen Jan vijf was, dacht hij dat de aarde plat was/is.
 when John five was, thought he that the earth flat was/is

However, in Russian, past can never be used in such embeddings; whether or not the belief is true does not influence the picture. The sentence (33) can only mean that Ivan's belief about the Earth's behaviour refers to the particular (past) time of his believing. In other words, the embedded past is still perceived as past with respect to the actual speech moment; it cannot be understood as having universal validity.

- (33) # Ivan polagal, čto Zemlja vraščalas' vokrug Solnca/Luny.
Ivan thought that Earth rotated around Sun/Moon

If there is a choice between a generic and a referential reading for the subject in the complement clause, the referential reading is the only one available:

- (34) a. Lošadi edjat ovës. (generic or referential)
Horses eat oats
b. Lošadi eli ovës. (referential only)
Horses ate oats
c. Ivan polagal, čto lošadi eli ovës. (referential only)
Ivan thought that horses ate oats

A similar observation applies to sentences like (36) and their Russian counterparts. The sentence (35) conversationally implies that Mary is no longer pregnant (unless the context specifically disfavours such implicature). However, (36) carries no such implication:

- (35) Mary was pregnant a month ago.
(36) A month ago John said that Mary was pregnant.

This is not the case in Russian. For the Russian (37), a simultaneous reading is not the most salient one, but it is still available. Under this reading, the sentence implies that Masha's pregnancy does not continue to the present moment, just as it would if the content of the complement clause was expressed in a root clause.

- (37) Mesjac nazad Ivan dumal, čto Maša byla beremenna.
Month ago Ivan thought that Masha was pregnant

Consider also English examples like (38) (James, 1986, p. 68):

- (38) She said her name was Sue.

This sentence is perfectly fine as an answer to the question "What is that girl's name?". The Russian counterpart of it sounds odd in such a context, implying that the girl's name could somehow be changed by now (the past-shifted interpretation is of course slightly odd as well, on pragmatic grounds, because it would involve the girl's account of some previous name of hers):

- (39) Ona skazala, čto ee zvali Maša.
 She said that her called Masha
 'She said that she was called Masha'

Another interesting case is presented by sentences like (40), uttered in a situation where I meet a friend in Utrecht on the day when I expected him or her to be away in Amsterdam:

- (40) I thought you were in Amsterdam!

Here the past is used to emphasize the fact that the speaker's opinion has turned out wrong at the present moment. In Russian, embedded Past does not express this implication. The reason for this is, again, that in the English case, the time of the person allegedly being in Amsterdam (expressed by *were*) can have "continuing applicability". In Russian (41), on the other hand, the speaker's opinion can only be valid for a past situation.

- (41) Ja dumala, čto ty byl v Amsterdame.
 I thought that you were in Amsterdam
 'I thought you was/had been in Amsterdam'.

All these examples show that the Russian simultaneous past is interpreted as "real" past with respect to utterance time, and when such an interpretation is disfavoured by the context, the sentences become odd. Finally, consider examples (42)-(44) from (Boeck, 1957, 1958). Boeck (1957, 1958) shows that simultaneous past, and not present, is strongly preferred in the following cases:

– when the past character of the action is somehow stressed or important in the discourse:

- (42) Vy zametili, kak on bral knigi?
 You noticed how he took books
 'Did you notice him taking (the) books?' (Boeck, 1957, p. 213)

– when the embedded clause contains a time adverbial referring to the past:

- (43) ...uznav, čto dvor naxodilsja/*naxoditsja **v to vremja**
 having-known that court was-located/*is-located in that time
 v Carskom Sele, rešilas' tut ostanovit'sja.
 in Tsarskoe Selo, she-decided here to-stay
 'Having found out that the (tsar's) court was **then** in Tsarskoe Selo, she decided to stay there'. (Boeck, 1957, p. 213)
- (44) Aleksej znal, kak dorogi *(byli) devuškam-soldatam vse eti
 Aleksey knew how dear *(were) girls-soldiers-DAT all these
 veščički **v te trudnye vremena.**
 little-things in those difficult times

'Aleksey knew how dear all these little things were to female soldiers **in those difficult times**'. (Boeck, 1958, p. 217)

The first example shows that past tense is preferred if the speaker wants to stress the fact that the embedded event was located at a particular past moment (for example, (42) could be pronounced by someone interrogating a witness of a particular instance of book theft). In the latter two examples, present tense is completely excluded unless we remove the adverbial, which "ties" the embedded event to a specific past moment.

To sum up, all the examples discussed above show that the Russian simultaneous past is evaluated with respect to the time of the actual speech situation, so it is perceived by speakers as past relative to the utterance moment. This is not the case in "SOT languages" such as English and Dutch.

It should be noted that several authors have commented on instances of simultaneous past interpretations in non-SOT languages. Sharvit (2003) claims that languages like Modern Greek occupy an intermediate position between SOT and non-SOT languages: in Modern Greek, both present and past can be embedded under a past verb to obtain a simultaneous reading. As for other non-SOT languages like Japanese or Hebrew, she notes that embedded past in these languages can sometimes have a simultaneous reading if the embedding verb is factive. The fact that in Japanese embedded past can receive a simultaneous reading under a factive matrix verb is also mentioned in (Ogihara, 1999b), as demonstrated by the example (45) (his (13a)):

- (45) Taroo-wa zibun-ga gan-dat-ta to sitte i-ta.
Taro-TOP self-NOM cancer-be-PAST that know-PAST
'Taro knew that he had cancer'

Ogihara claims that this is an exceptional case that does not alter the general picture (namely, past under past always encodes precedence in Japanese complements). He suggests that, since the complements of factive verbs are presupposed to be true, they can be analysed differently from complements of non-factives: the complement of the factive verb is in fact a kind of NP meaning "the fact that... ", so it is scoped out and does not fall within the scope of the matrix tense. (In fact, the NP analysis is a variation of *de re* analysis, as Ogihara (1996) presents it. There, too, the embedded T is supposed to move out of the scope of the matrix tense.)

This is an interesting fact because it is in line with the findings for Russian, although it seems that this kind of phenomenon is much more limited in Japanese than it is in Russian. In particular, Ogihara claims that the Japanese sentence (45) loses its simultaneous interpretation if the factive verb 'know' is replaced by a verb like 'believe'. However, it is an important observation that such cases are also present in other "non-SOT" languages.

Japanese simultaneous past examples are also discussed in (Kusumoto, 1999). She gives examples similar to the Russian (43)-(44), where embedded past tense in a complement clause can have a simultaneous meaning if the clause contains a temporal adverb anaphoric to the matrix tense:

- (46) Junko-wa [Satoshi-ga sonotoki byookidatta to] itta
 J-top S-nom that-time sick-past comp say-past
 'Junko said that Satoshi was sick then'

If the adverb *sonotoki* 'then' is understood to mean the time of Junko's saying, the embedded past tense is also simultaneous with this time. Kusumoto suggests that the vacuous past tense is licenced by the temporal adverbial: it introduces a *PAST* operator, which in her system is responsible for licencing a (vacuous) *past* morpheme. However, in Russian the cases of simultaneous past are not limited to clauses with a temporal adverb. This is does not seem to be the case in Japanese, either, as Ogihara's example (45) shows, so Kusumoto's analysis does not seem sufficient.

Free Indirect Discourse in Russian

In the preceding section, I have shown that the interpretation of simultaneous past in Russian is different from that in English, or, more accurately, that not all interpretations of simultaneous past that are present in English are available in Russian. A similar picture arises from the analysis of Free Indirect Discourse in the two languages. Free Indirect Discourse (FID) is a type of narrative discourse where the narrator renders the thoughts of a character without resorting to direct quotation or indirect speech. The formal properties of FID are use of temporal and spatial adverbials and demonstratives as in direct speech, and use of tense and personal pronouns as in indirect speech. An instance of English FID is given in (47): the adverb *tomorrow* would normally be excluded in indirect speech, but in FID it can be used, although the choice of tense (past) indicates that this is not a direct quotation:

- (47) Tomorrow was Monday, Monday, the beginning of another school week! (D.H. Lawrence, *Women in Love*) (Doron, 1991, p. 53, (2))

Thus, in English, past tense is used in FID which is part of a past narrative to denote the present of the character whose thoughts are being rendered by the FID. As for Russian, it has been claimed to differ from English in this respect: Russian uses present tense, not past, to describe the "now" of the character in a FID context included in a past narrative (see e.g. Babynyshev & Matushansky, 2006). This is consistent with the fact that Russian normally uses present in indirect speech to denote the now of the matrix clause speaker. However, Russian FID can also use past imperfective in describing events simultaneous with the now of the character.

Hence, the use of tenses in FID is fully parallel to their use in Russian indirect speech (in particular, in complements embedded under past). The narrative in (48) can serve as an illustration; in the translation, I mark the tense of the verbs used in the FID context.

- (48) Ona sprosila sebja: xorošo li ona postupila, što otkazala čeloveku tol'ko potomu, što ej ne nraivtsja ego naružnost'? Pravda, èto neljubimyj čelovek i vyjti za nego značilo by prostit'sja navsegda so svoimi mečtami, svoimi ponjatijami o sčast'e i supružeskoj žizni, no vstretit li ona kogda-nibud' togo, o kom mečtala, i poljubit li? Ej uže 21 god. Ženixov v gorode net. Ona predstavila sebe vsëx znakomyx mužčin – činovnikov, pedagogov, oficerov, i odni iz nix byli uže ženaty i ix semejnaja žizn' poražala svoeju pustotoj i skukoj, drugie byli neinteresny, bescvetny, neumny, beznравstvenny. Laptev že, kak by ni bylo, moskvič, končil v universitete, govorit po-francuzski; on živet v stolice, gde mnogo umnyx, blagorodnyx, zamečatel'nyx ljudej, gde šumno, prekrasnye teatry, muzykal'nye večera, prevosxodnye portnixi, konditerskie... (A.P. Chekhov. *Tri goda*).

'She asked herself: did she do a good thing when she refused the proposal of a man just because she didn't like his appearance? True, it was(PRES) an unloved man, and to marry him would mean to say farewell to her dreams, her ideas of happiness and of married life, but would she ever meet the one she had dreamed of, and would she fall in love with him? She was(PRES) already 21. There were(PRES) no potential fiancés in the town. She imagined all her male acquaintances – officials, teachers, military officers, and some of them were(PAST) already married and their family life struck(PAST) one as empty and boring, others were(PAST) uninteresting, colourless, unintelligent, immoral. And Laptev, in any case, was(PRES) from Moscow, he had been to university, he spoke(PRES) French; he lived(PRES) in the capital where there were(PRES) many intelligent, noble, remarkable people, where it was(PRES) noisy, where there were(PRES) wonderful theatres, musical soirées, excellent tailors, confectioner's shops...'

Here in one paragraph we see alternating present and past forms. Earlier in this section, I have shown that in Russian complement clauses, simultaneous past appears when the events are viewed from the (actual) speaker's perspective. The same reasoning can be applied to FID: past appears when the narrator's point of view takes precedence over the character's point of view. This effect is analysed in great detail in (Padučeva, 1996). According to Padučeva, FID involves two "voices": that of the narrator and that of the character (the subject of speech/thought represented in FID). The use of present or past imperfective tense reflects the predominance of one over the other: if the character is in the foreground, present is used; if the

narrator's perspective is taken, past is used. Although I am not committing myself to the technicalities of Padučeva's analysis, I think that her intuition is correct. Thus, it seems that the same rules that govern the choice of past or present in past-embedded complements work for FID contexts: when the narrator or speaker's point of view takes over, past is used; otherwise, present is preferred.

4.1.3 Summary

To sum up, the data from Russian presented above show that in case of simultaneous past, the eventuality denoted by the embedded verb is regarded as past from the point of view of the (actual) speaker. This is correctly predicted by the *de re* interpretation theory: the speaker is referring to some actual past moment which is substituted for another moment in the belief of the matrix subject. However, not all the English data fall into this pattern: English simultaneous past does not necessarily denote an eventuality located in the past from the point of view of the actual speaker; rather, the point of view shifts to the subject of the embedding clause. Hence, a different mechanism is needed to account for the English simultaneous past.

In the next section, I will develop my account of the simultaneous past in SOT languages like English and Dutch. It will be based on a binding relation between the two tense morphemes, implemented via an agreement relation between the instances of the T(ense) feature.

4.2 SOT: The account

In this section, I explain how the simultaneous past interpretation is generated in English, and why it is absent in a non-SOT language like Russian. The essence of the proposal is that there is a binding relation between the embedded and the embedding T heads in English complements, while no such relation can be formed in Russian, for the following reason: the relation in question is based on an Agree mechanism between the T features present in both clauses; a chain of agreeing feature instances proceeds via the embedded C head, which carries an instance of T feature; however, in Russian, C lacks a T feature.

4.2.1 SOT in “SOT languages”

The account of SOT that I propose inherits its basic assumptions from the approach of (Enç, 1987, 2004). She argues that the simultaneous interpretation of embedded past is caused by a binding relation between tense morphemes. As noted in (Khomitsevich & Reuland, 2007), Enç implements what she qualifies as binding by head-head dependencies: the relation is

established between T and T or T and C. This is not quite in the spirit of the traditional BT that she was presupposing. However, the current theory deals precisely with such relations where heads as probes play a crucial role in establishing the syntactic component in anaphor binding (Reuland, 2001, 2005, see also Chomsky, 2005). Pesetsky and Torrego (2001, 2004) develop an explicit model for representing feature based dependency chains in the C-T domain. I will rely on these approaches in developing a theory of temporal dependencies.

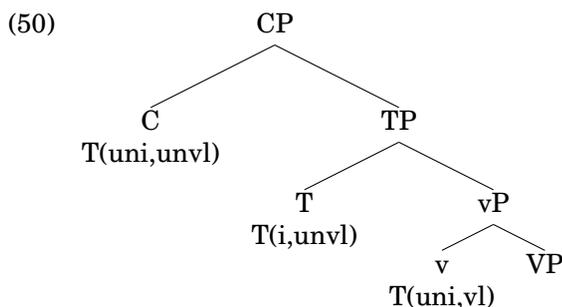
Let me recapitulate the SOT data that are in need of explanation. Consider again the English sentence given in (49).

(49) John believed that Mary was pregnant.

(49) can have two interpretations: one is a “past-shifted reading”, where the time of Mary’s supposed pregnancy precedes the time of John’s believing in it; the second is a “simultaneous reading”, where the times of these two events overlap. The latter is the reading that I will be concerned with here.

Based on the conclusions in the preceding sections, I argue that in sentences like (49), there is a certain relation established between the higher and the lower tense when the simultaneous reading of the embedded past is obtained. I propose to define this relation in structural terms: it holds in complement clauses only, and is blocked in other types of embeddings. This relation can be defined in terms of feature agreement, in the sense of (Pesetsky & Torrego, 2004b) and (Reuland, 2005).

I adopt PT’s (2004b) system, in which the tense feature on T is unvalued but interpretable, and the tense feature on *v* is valued but uninterpretable. It is also important for my approach that C bears an instance of the T feature; I will assume that it is unvalued and uninterpretable. In the course of the derivation, the unvalued tense feature on T probes and finds a valued instance of the tense feature on *v*, agrees with it and gets valued. Further, the unvalued tense feature on C probes and finds a valued tense feature on the T head; I am abstracting away now from the multiple ways to value the tense feature on C discussed in (Pesetsky & Torrego, 2001), and assume that in all cases, the value of the tense feature on C eventually comes from the T head, which in turn had received it from *v*. This results in the clause structure in (50) (repeated from (21) on page 36; (un)*i* stands for (un)interpretable, (un)*v*_I for (un)valued):

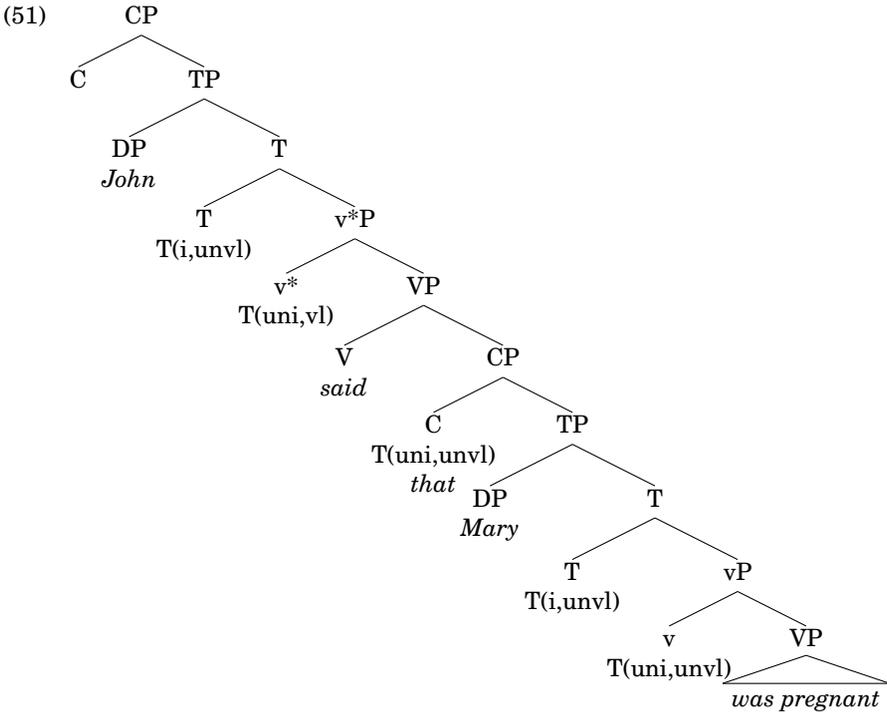


I propose that in case of the simultaneous reading of the past tense in complement clauses, the embedded verb enters the derivation with its tense feature unvalued. It then gets its value from the higher verb in the embedding clause. This is a result of an agreement relation established between the higher and the lower verbs. In what follows, I explain in detail how this happens.

Recall that in PT's system, an Agree relation can also be established between two unvalued features. I propose that this is what occurs in this case: an Agree relation is established between the tense feature instances on *v*, *T* and *C*, although none of these instances is valued. Thus, as the derivation reaches the (embedded) CP phase, the features on *T* and *v* are still unvalued. This could of course lead to a derivation crash. Importantly, however, in PT's framework several features linked by Agree operations become instances of one and the same feature. We can therefore assume that since one instance of the unvalued tense feature is found on the phase edge – on *C*, which is accessible to further syntactic operations, all the instances of this feature remain accessible and can survive into the next phase.³

At the next stage, the complement CP is merged with the higher VP. It is then attracted by the *v* head to value *v*'s object agreement ϕ -features. Thus, an agree relation between *v* and *C* is established. Both matrix *v* and *C* carry a tense feature, which is unvalued on *C* but valued on the matrix *v*. Following (Reuland, 2005), I argue that the ϕ -feature dependency can extend to a T-feature dependency, so the T-features on *v* and *C* enter into an Agree relation. At this point, all instances of *T* in the complement clause are valued; the same value is later obtained by the matrix *T*. The resulting structure, for a sentence like (52), is given in (51) (I ignore the instances of the tense feature on the matrix *C* and *V*, as well as the subject DPs, for the sake of simplicity):

³Note that this modification is not compatible with the strictest interpretation of phase theory; however, this is a very general problem that also arises in theories of anaphora resolution: although a phase is supposed to be sent off to the interfaces and made unavailable to the further computation, pronominal elements inside it still have to be accessible if their interpretation depends on elements in higher phases.



(52) John said that Mary was pregnant

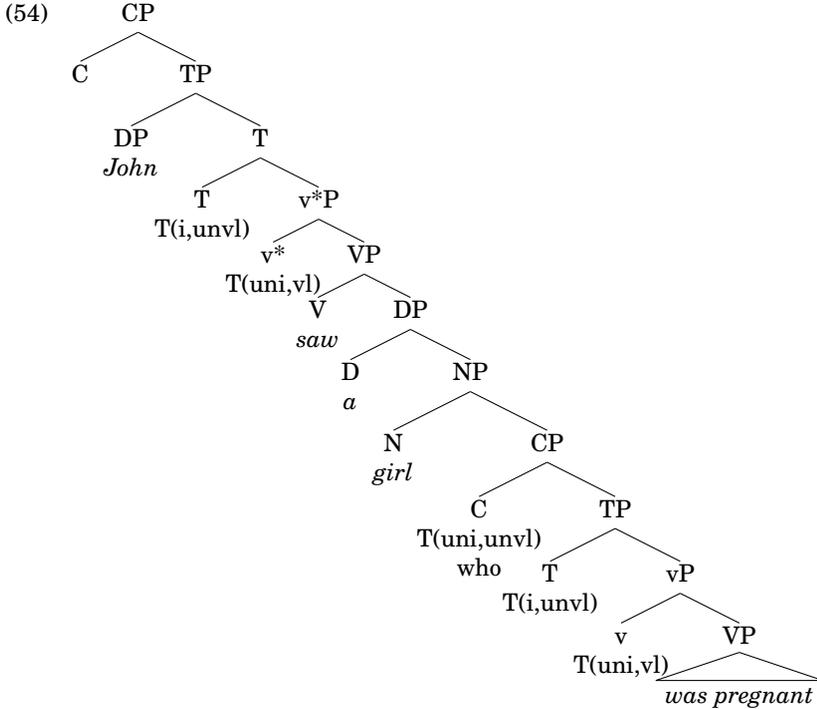
In this sentence, the matrix and the embedded tense features on T share the same value. If the unvalued verb in the complement clause is embedded under another past verb, it will receive the value *past*. Consequently, the two verbs share the same temporal interpretation. Of course, the fact that two verbs both bear the feature *past* does not by itself mean that they denote simultaneous (overlapping) activities; what is important is that here they bear instances of *the same* feature. Following (Reuland, 2005), I propose that the feature agreement relation is interpreted as a binding relation at LF, which means that the two activities denoted by the verbs are simultaneous.

The way the theory stands now, the agreement process in question can only be formed in complement clauses. Consider a relative clause, such as (53):

(53) John saw a girl who was pregnant.

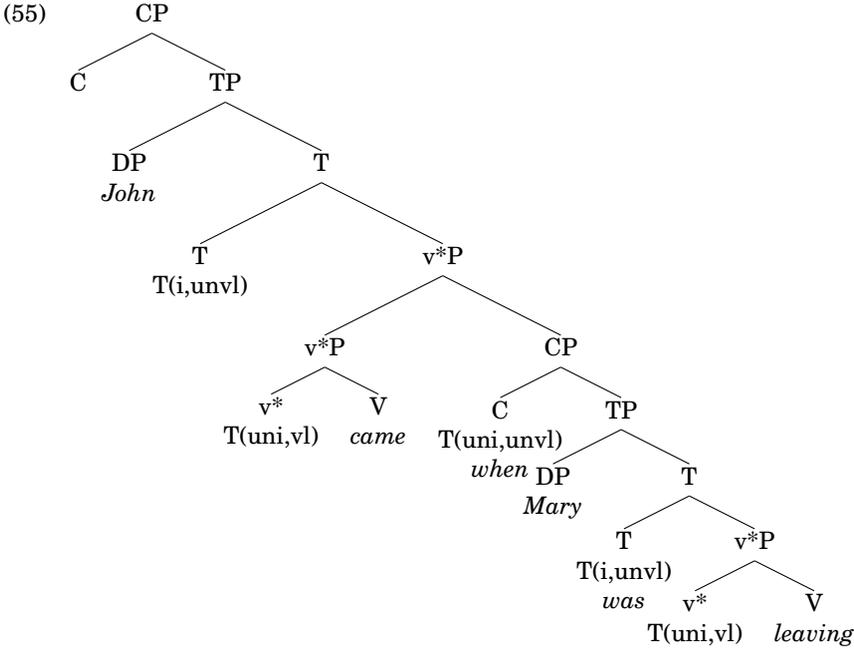
Suppose the tense feature on the embedded verb enters the derivation unvalued, so that by the time of completion of the CP, the C bears an unvalued tense feature. The next step is to merge the relative clause to its head, the DP *a girl*, which in turn is merged to the matrix V (I am leaving aside the details of relative clause formation). Consequently, there is

no Agree relation established between the matrix *v* and the embedded *C*, since the ϕ -features on the matrix *v* are valued by the object DP *a girl*. The *T* feature on the embedded verb cannot receive any value, and the derivation crashes. So the “bound” reading is not available in relative clauses. The relevant structure is given in (54) (the *T* feature on the embedded *v* is marked as valued, since the derivation will crash otherwise).



Note that a sentence like (53) does have a meaning where the two times introduced by the verbs overlap (a “simultaneous” reading). However, our analysis shows (contra Ojihara (1996) and Enç (1987)) that it is not derived in the same way as the simultaneous reading of complement tenses – it is not a result of a binding relation. Rather, the simultaneous interpretation is a result of “accidental” overlap between the matrix and the embedded tense. (An analogy with covaluation of pronouns can be drawn here.)

In case of adjunct clauses, an embedded *C* does not establish a link with the embedding *v* via ϕ -features, since the embedded clause is not an object and does not enter into the process of object agreement. Thus, no “bound” reading is available, and the temporal relations between the eventualities described by the verbs are determined by other factors, such as the semantics of the complementizer. The relevant structure for the sentence (56) is given in (55).



(56) John came when Mary was leaving.

To sum up, I have argued for an analysis of temporal dependencies in SOT configurations as Agree chains between heads. This type of approach is in line with several recent accounts. As discussed in section 3.2.2, a similar mechanism is applied to pronouns by Kratzer (2006). (Unlike in her earlier work (Kratzer, 1998), (Kratzer, 2006) does not extend her analysis to tenses.) A crucial difference between the two analyses is that Kratzer has to postulate a feature transmission mechanism which can “go over” heads that do not display the relevant features present on the head and tail of the chain. The analysis that I propose takes into account the feature makeup of all intermediate steps.

Another similar approach is presented in (Manzini, 1994), who proposes an analysis of subjunctive in terms of a head-head dependency. According to (Manzini, 1994), subjunctive in Romance is “an indefinite T bound by an intensional operator in a syntactic dependency” (p. 241). In particular, subjunctive is selected by sentential operators Neg (negation), Q/Wh (question), If (if-clause); importantly, both T and the operators are heads. That this is indeed a syntactic dependency is shown by the fact that it is sensitive to islands:

(57) E' andato perchè è/*sia stanco?

‘Has he gone because he is/is:subj tired?’ (Manzini, 1994, p. 244, (9))

In (57), the operator Q cannot license the subjunctive inside the adjunct clause introduced by *perchè* “because”. Manzini (1994) suggests that in case of Neg and other such operators, “[t]his dependency is roughly of the form (C, I, V, C, I), if we take the higher C to represent the position of the operator and the lower I the position of the embedded subjunctive” (p. 245). However, Manzini (1994) does not present a precise syntactic implementation of this idea.

What remains to be mentioned is the problem of forward-shifted readings. I have proposed two possible ways of deriving a simultaneous past reading: one is a *de re* interpretation mechanism, the other is an Agree chain mechanism. So far, neither of them excludes a forward-shifted reading. As discussed extensively by Abusch (1997), the “independent” tense theory (which produces the *de re* interpretation) cannot alone explain the absence of such readings. As for the Agree mechanism, it does not say anything about the possible interpretations of the embedded tense once the chain is not formed. I will follow Abusch (1997) in this respect and assume the existence of an Upper Limit Constraint (ULC), which establishes the now of the intensional predicate (the matrix time) as the “upper limit” for the interpretation of an embedded tense (see section 3.2.2). This is an independent constraint which is always valid for tenses in the scope of an attitude operator.

There is one more detail to my account that I should comment on. Once we allow temporal features of verbs to enter the derivation unvalued and receive their values later, a question arises what happens if a *matrix* verb is generated with an unvalued T feature. So far, nothing in my theory prohibits an unvalued matrix tense feature to probe down, find a valued tense feature on the embedded C/T, and get its value from it. Thus, in (52), *said* would get its past feature from *was*.

Of course, this is very counterintuitive, since it is normally assumed that matrix tenses are always evaluated with respect to speech time; it is impossible for them to depend on their embeddings for their interpretation. If there are multiple complements, then lower clauses is supposed depend on higher ones, but never the other way around.

It is easy to demonstrate that such a derivation is indeed impossible. In section 3.1.1, it was demonstrated that verbs which allow a simultaneous past reading when they are embedded are always stative verbs – verbs that denote states and not events. If an eventive verb is embedded in a complement clause, it will always get a past-shifted interpretation, even if the embedding verb is a stative one:

(58) Mary believed that John came.

In principle, nothing prevents the time of John’s coming to be simultaneous with Mary’s thinking: imagine that Mary sees John entering the room and at this very moment forms an opinion about that. However, it is not a licit reading of (58); if we want to get it, we are forced to use the continuous

form *was coming*. On the other hand, the embedding verb can be eventive, like *said* in (52), repeated here as (59):

- (59) John said that Mary was pregnant.

There is no prohibition for *was pregnant* to be interpreted as simultaneous to *said*. The contrast between (58) and (59) indicates that the dependency between the two tenses is asymmetric.

In fact, a similar issue arises in connection with the pronominal theory of (Reuland, 2005). There, nothing in the agreement relation itself prevents the anaphor appear in the subject position and be bound by the direct object. However, this derivation is ruled out on independent grounds. It seems that my temporal theory is in need of a Condition C-like constraint on temporal binding.

This constraint can be implemented along the lines of (Reinhart, 2000), who shows that binding is impossible in an example like (60a), since it would yield the uninterpretable (60b): “*Lucie* is not the type of object that can be bound by the λ -operator whose sister is *she* (since it is not a free variable)” (Reinhart, 2000, p. 10).

- (60) a. She said we should invite *Lucie*.
 b. She (λx (x said we should invite *Lucie*)) (Reinhart, 2000, p. 9, (16a-b))

Since the feature-sharing relation between the two T nodes is interpreted as a binding relation, this reasoning can be applied to it as well: the T with a valued T-feature simply cannot be bound.

This discussion raises the question which has not been addressed so far (either here or in (Reuland, 2001, 2005)), which is how the Agree relations translate into binding. This could be done along the lines of (Řezáč, 2004), who develops a strategy to ensure that “ α Merged in [Spec, HP] λ -binds only the goal Γ of Agree by H” (Řezáč, 2004, p. 1). This allows him to account for the so-called Copy Raising constructions. Copy Raising is illustrated in (61a); it is shown to be impossible in (61b).

- (61) a. John seems like he is tired.
 b. * John seems that he is tired.

Řezáč proposes that an Agree relation between the DPs *John* and *him* is blocked if the intervening complementizer bears lexically specified default ϕ -features, which is the case with *that*, and is licit otherwise (in case of *like*). This mechanism differs from the one that is proposed in this thesis, in that the complementizer is supposed to block an Agree relation and not transmit it. However, in fact there is no conflict between the two approaches. In my model, the complementizer only transmits the Agree relation if its T feature is not valued within its clause. Otherwise, no Agree relation takes place.

4.2.2 SOT in Russian

The analysis that I developed in the previous section shows how a simultaneous reading is derived in English. Now I can distinguish effectively between SOT-languages like English and Dutch and a non-SOT language like Russian. Let me recapitulate the relevant differences between the two language types here. Examples (17a)–(17b) on page 56, section 3.1.5, repeated here as (62a)–(62b), show that past embedded under past in Russian complement clauses does not normally have a simultaneous interpretation; it is usually interpreted as past-shifted. To denote the simultaneous interpretation, present is used. These are the data usually discussed in the literature on SOT in Russian.

- (62) a. Ivan skazal, čto Maša bolela.
 Ivan said that Masha ailed
 ‘Ivan said that Masha had been ill’
- b. Ivan skazal, čto Maša boleet.
 Ivan said that Masha ails
 ‘Ivan said that Masha was ill (at the time of saying)’

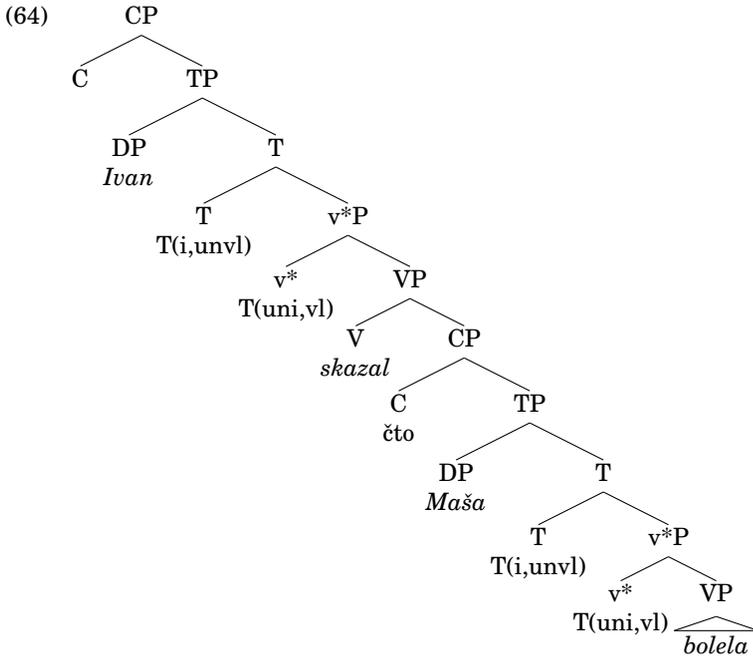
In section 4.1.2, I showed that cases of simultaneous past do exist in Russian, and that they are in fact quite widespread. Sentences like (63) (repeated from (21), page 94) are the case at hand:

- (63) Ona bespokojno šarila vokrug sebja rukami, i on
 she anxiously fumbled around self with-hands, and he
 dogadalsja, čto ona iskala ubežavšuju Murku, v kotoroj
 guessed that she searched-for run-away Murka in whom
 nuždalas’ vsjakij raz, kogda popadala v zatrudnitel’noe
 needed every time when got in difficult
 položenie.
 situation
 ‘She was fumbling around anxiously with her hands, and he guessed that she was searching/*had been searching for the run-away Murka [a cat’s name], who she needed every time when she was in a difficult situation’. (simultaneous reading)

However, in section 4.1.2 I argued at length that the simultaneous interpretation of past in Russian complement clauses is not analogous to English simultaneous past. I demonstrated that the past tense verb in the embedded clause is still perceived as past with respect to the utterance time, while in English it is not necessarily so; for instance, “general truth” statements whose validity extends to the utterance time can be reported using simultaneous past in English, but not in Russian (see examples (31)–(33)). These facts led me to the conclusion that the mechanisms of obtaining the simultaneous past interpretation are different in Russian and English.

In the previous sections, I proposed that the simultaneous past in Russian is interpreted *de re*. In English and Dutch, on the other hand, there is an agreement mechanism connecting the T(ense) features of the two predicates, which is interpreted as a binding relation. As a result, we get the specific “SOT reading” which is the main subject of debate in the literature on SOT.

The question is where the difference between Russian and English stems from. I propose that the problem lies in the feature composition of the C node. While in English, the C head bears an instance of a tense feature, the Russian C does not. For this reason, in Russian the embedded verb cannot enter the derivation with an unvalued tense feature: there is no way it could get a value from the higher verb. The distribution of T(ense) features is demonstrated in (64), the structure of (65):



- (65) Ivan skazal, čto Maša bolela.
Ivan said that Masha was-ill

Suppose the T feature on the embedded verb is unvalued. In that case, the embedded T would probe for the T feature on *v* but no valuation would ensue. However, the agreement process stops here, since C does not carry a T feature in need of valuation and so no link between C and T is formed. The lower T feature would thus have to remain unvalued, and the derivation would necessarily crash. So, the bound reading of embedded past is unavailable in Russian.

Thus, there are two readings for past-under-past in Russian complement clauses: it can be either past-shifted, or simultaneous *de re*. I turn to the derivation of English and Russian past-shifted readings in the next section. As for the simultaneous reading, it can also be expressed by using embedded present, which unambiguously denotes simultaneity (which is probably why present is often preferred). Russian present under past will be discussed in section 4.3.2.

Finally, I turn to the implications of my hypothesis about the Russian C. Recall that according to Phase theory, C and T are closely linked: C is the phase head, and T gets its features derivatively from it. A technical implementation of this idea is Pesetsky & Torrego's proposal that a tense feature is present on C. Consequently, the lack of a temporal feature on C indicates a different relation between the C and T heads. In Chapter 5 (section 5.3.1), I will show that this is indeed the case with Russian: in this language, T is a phase head and not C.

4.3 Remaining issues

4.3.1 Deriving past-shifted readings

So far, I have concentrated on deriving the simultaneous reading of embedded past – the topic that most of the works on SOT focus on. Now I turn to the other type of temporal interpretation found in SOT environment, namely, the past-shifted readings. Such a reading is possible in (66b) and obligatory in (66a):

- (66) a. John said/believed that Mary won the race.
 b. John said/believed that Mary was pregnant.

I mentioned in section 3.1.1 that past-shifted readings are possible (although, for some speakers, not quite natural or salient) with stative embedded verbs, but obligatory with eventive embedded verbs (as in (66a)). It does not matter whether the embedding verb is stative or eventive; in any case, the time interval to which the embedded eventive verb refers to will precede the matrix time. In my discussion of unvalued tense features, I did not mention this problem: apparently, temporal binding can never hold for eventive verbs. In fact, this question holds for most SOT theories, which are limited to stative verbs and leave eventive verbs outside the scope of the discussion. I will assume that temporal features on embedded eventive verbs can in fact be unvalued and receive their past value from the matrix T; however, the specific semantics of eventive verbs forces the precedence interpretation anyway. I conclude this from the fact that simultaneous interpretation is possible for eventive verbs under special circumstances – when the embedding verb is a perception verb or has a meaning of belief based on perception, as in examples like (17) and (18b), repeated here as (67a) and (67b):

- (67) a. Elliott observed/noticed/perceived that Josephine got hurt.
 b. He didn't realize that his car hit the curb.

In any case, the question arises how to derive past-shifted readings of past-under-past. There are two ways to do that. One possibility is that the embedded past is evaluated as past with respect to utterance time, but is ordered as preceding the matrix time. The other possibility is that it is ordered as past with respect to the matrix time, and not to the utterance time; this is the derivation assumed in the operator theory of tenses.

In fact, the first option is fully compatible with my approach. I argue that (embedded) tenses can be interpreted *de re*; this automatically gives us two possible readings – simultaneous (if the embedded tense is interpreted as anaphoric to the matrix tense) and past-shifted (if it refers to some time interval preceding the matrix time); but not the forward-shifted reading, due to the Upper Limit Constraint.

However, having only this option would mean that “past-shifted past” is always interpreted *de re*. Kusumoto (1999) argues that it is not always desirable to have a *de re* interpretation for embedded past. In particular, the following example seems convincing:

Suppose that on July 2nd, John said, “Anita went to NY yesterday.” Suppose further that John had no idea when on July 1st Anita went to NY. On July 3rd, we can report what he said as “Yesterday John said that Anita went to NY the day before (yesterday)”. In this scenario, what my utterance should mean is that John said of the day before yesterday that it has a property of containing the time of Anita's going to NY, not that John said of a particular time in the day before yesterday that it has a property of Anita's going to NY. If we interpret the embedded past tense in the example above as *de re*, what we get is the latter interpretation (Kusumoto, 1999, p. 229).

I conclude that embedded past can also be evaluated as past with respect to the time interval of its embedding tense. In that case, the semantics of past can be taken to denote precedence to some evaluation point; this point can be either the utterance time or some other time introduced in the discourse.

However, not any reference point will do. Normally, in a discourse, an eventive verb in the past tense will denote an action following the action described by a preceding verb, so that it “moves forward” the narrative (see Partee, 1973, Borik, 2002, a.m.o). Consider (68):

- (68) a. John came into a shop. He bought a newspaper.
 b. John came into a shop. He had bought a newspaper.

(68a) is understood as meaning that John bought the newspaper at the shop mentioned. It cannot mean that he bought it previously at a newsstand outside the shop. To convey this meaning, we have to use pluperfect, as in (68b).

The situation is different when the verb is embedded in an attitude context. There, as the examples in (66) show, past easily gets a past-shifted reading. This is in line with Schlenker's (2003) approach to tenses as shiftable or unshiftable indexicals, outlined in section 3.2.2. Babyonyshev & Matushansky (2006) expand his theory of present tenses to cover past as well; they further argue that in Russian complements "the past tense can be indexical and indicate temporal precedence with respect to the time of utterance; the availability of the simultaneous interpretation for the past tense is conditioned by the properties of the matrix verb, but in principle possible; it can also be shifted, in which case it indicates temporal precedence with respect to the matrix predicate" (p. 11-12). This analysis can be applied to the English past as well. So, I conclude that both English and Russian past can be "shifted" so that the eventuality denoted by the embedded past verb is evaluated as preceding the matrix time.

4.3.2 Present under past in Russian and English

The facts I deal with in this section are the following: the embedded present in the English (69a) can only have a "double-access" reading, while in the Russian (69b) it can easily have a simultaneous interpretation.

- (69) a. John said that Mary is pregnant.
 b. Ivan skazal, čto Maša boleet.
 Ivan said that Masha ails
 'Ivan said that Masha was ill (at the time of saying)'

In this section, I will develop and comment upon Schlenker's (1999 et seq.) treatment of present tense in Russian and English. In section 3.2.2 I summarized Schlenker's theory of present tenses: he treats them as indexicals which always refer to the "now" of some speech/thought context. The difference in the behaviour of the present tense in English and Russian is explained as follows: English present tense is an unshiftable indexical, which always refers to the now of the actual speech act; Russian present tense is a shiftable indexical, which means that it can be "shifted" to denote the now of a reported speech or thought act.

The obligatoriness of the double-access reading in English then follows from Abusch's theory of double-access interpretation developed further in (Schlenker, 2002, 2003). Recall that Abusch's Upper Limit Condition prohibits tenses embedded in belief contexts to refer to times later than the belief time. Therefore, the time denoted by the embedded verb in (69a)

has to overlap the speaking time. But it also has to overlap with the utterance time since it cannot be “shifted” from it. Hence the double-access semantics.

Russian, on the other hand, does not impose the latter condition on the interpretation of the embedded present, so it can be understood as referring only to the reported context and not to the actual context (the double-access reading is also available; however, it does not seem necessary to treat it as a separate reading in the Russian case: primarily, the embedded present is valued with respect to the matrix time, and in principle it can also overlap with utterance time).

As noted in section 3.2.2, there is a potential problem for Schlenker’s account noted by (Babyonyshev & Matushansky, 2006), illustrated in (42) on page 72: adverbial expressions denoting attitudes, such as “according to Masha’s claim”, do not license shifting of the Russian present tense. In section 3.2.2 I argue that the adverbials like the one in (42c) should not be analysed as proper attitude reports, contrary to what BM assume. Let me provide another argument to demonstrate that we are not dealing with a case of local syntactic licensing or binding here. As I showed in section 1.4.2, complement clauses, as well as some other types of embedded clauses in Russian, can combine with a “pronominal correlative” in the matrix clause. Thus, the clause in (70b) is no longer a complement of the attitude verb. However, introducing the correlative does not change the temporal interpretation:

- (70) a. Ivan znal, čto Maša boleet.
Ivan knew that Masha is-ill
'Ivan knew that Masha was ill'
- b. Ivan znal o tom, čto Maša boleet.
Ivan knew about that that Masha is-ill
'Ivan knew that Masha was ill'

The tense options for sentences (70a) and (70b) are exactly the same, although the syntactic structure is different. In both cases, embedded present denotes simultaneity. I conclude that the analysis of Russian present proposed in (Schlenker, 1999, et seq.) and (Babyonyshev & Matushansky, 2006) is on the right track.

4.3.3 Complements of NPs

Not only verbs, but also NPs can induce SOT effects in their complements. Recall the example (3a) from section 3.1.1, repeated here as (71):

- (71) Mary’s announcement that she was pregnant is irrelevant now.

This sentence can be understood as stating that Mary earlier claimed to be pregnant (at the moment of announcement). This is an instance of a

simultaneous past; however, there is no overt past verb to licence it, because the embedding element is an NP and not a VP. For my account, it means that, at the first sight, there is no suitable tense feature to bind the embedded past tense.

Indeed, Abusch (1997) argues that in sentences like (71) there is no syntactic mechanism that could trigger SOT effects. However, there is more evidence in the literature that NPs can have an independent temporal status. (Enç, 1986) argues in favour of this idea, citing examples like (72):

- (72) John will meet every hostage at the president's party. (Enç, 1986, p. 409, (12))

This sentence does not mean that the people that John is going to meet will actually be hostages at the moment of the party; it means that John will meet the people who were hostages at some past moment. On the basis of this and similar examples Enç concludes that the temporal reference of NPs does not have to depend on the tense of the predicate of the clause.

Musan (1999) demonstrates that the picture is much finer. In fact, some NPs obligatory depend on their predicates for temporal interpretation, and some do not. Musan argues that this follows from the distinction between quantification over individuals and stages of individuals: "(...) quantification over individuals triggers temporal independence, whereas quantification over stages triggers temporal dependence" (Musan, 1999, p. 658).

What I would tentatively propose here is that NPs carry a temporal feature which in the default case is set to the value of the local T, along the lines of Pesetsky & Torrego. (Note that in PT's model only structural case is represented as an instance of tense, however we would need to endow all NPs with an ability for temporal reference). However, certain types of NPs (for instance, attitude NPs like *announcement*) can somehow override this default value. That would solve the problem with SOT environments like (71).

Of course, this would have consequences for our treatment of relative clauses. Recall the structure in (54); if we suggest that there is an uninterpretable tense feature on the NP *girl* which gets its value from the matrix T, then in principle the embedded verb should be allowed to carry an unvalued tense feature which gets the same value from the matrix verb via the head NP *girl*. This would mean that there is a bound simultaneous reading for relative clauses – something that should not bother us since this reading would in any case be indistinguishable from the "covaluation" simultaneous reading.

An alternative would be to propose that embedded tense in cases like (71) is always evaluated with respect to the utterance moment and interpreted *de re*. Consider the way the interpretation of (71) proceeds. The sentence is in fact ambiguous: the NP *announcement* can denote either an announcement made earlier, or a present announcement (referring to an earlier pregnancy). However, if we understand the announcement as

simultaneous with the alleged pregnancy, the tense of the embedded verb tells us unambiguously that the announcement was also made in the past. This is precisely because the noun has no temporal marking of its own, so the interpretation of the embedded past has to depend on the utterance time.

4.3.4 Embeddings under future

In my discussion of SOT, I was concerned with complements and other embedded clauses dependent on matrix clauses with a past predicate. However, as I mentioned in section 3.1.4, clauses embedded under future verbs also present several (less well-explored) problems. Although I do not have a definitive solution for these problems, I consider it necessary to highlight a number of possible analyses.

The first problem concerns complement clauses embedded under future. As shown in sentence (10b) on page 54, repeated here as (73b), English present tense can be “shifted” when it is embedded under future: in (73b), the time of Mary’s pregnancy does not overlap with utterance time. In contrast, present embedded under past has to overlap with the moment of utterance, thus producing the so-called “double access” reading, as in (73c).

- (73) a. John will say in 2010 that Mary was pregnant in 2009.
 b. In 2010, John will say that Mary is pregnant.
 c. John said that Mary is pregnant.

Besides, some authors claim that a dependent reading of both past and present is also available in relative clauses embedded under future verbs. Thus, Enç (1996) argues that in (74a), the time of the man’s speaking does not have to overlap with the moment of utterance: it can be evaluated with respect to Mary’s seeing the man. A similar example with present is given in (12a), and a dependent reading of past is demonstrated in (12b) and (12c) (p. 54, repeated below as (74b)–(74d)).

- (74) a. Mary will see the man who is speaking (Enç, 1996, p. 352, (27)).
 b. John will buy a fish that is alive (Ogihara, 1999a, ((13), p. 229)).
 c. I will charge you whatever time it took (Heim, 1994, p. 158, (42)).
 d. At the end of next term, I will give automatic A’s to all students who turned in their term papers on time (Abusch, 1998, p. 13, (1)).

However, it seems that the availability of such readings is context-dependent. Thus, according to (Stowell, 1995), there is no relative reading of the embedded tenses in (11) on page 54, repeated here as (75):

- (75) a. Adam will give an ice-cream cone to a boy who was sitting outside.
 b. Adam will give an ice-cream cone to a boy who is sitting outside.
 (Stowell, 1995, (7a,b), p. 385)

To be precise, Stowell (1995) considers that a dependent reading of past under future is marginally possible, especially if forced by adding an appropriate adverbial expression, such as *a few minutes before* in (76). However, he concludes that this reading is very hard to obtain and can be neglected.

- (76) Adam will give an ice-cream cone to a boy who was sitting outside a few minutes before. (Stowell, 1995, (8))

Nevertheless, Enç (1996) and Abusch (1997, 1998) base their accounts of the future tense – or, rather, of the English future auxiliary *will* – on the assumption that it licences relative temporal readings in all kinds of embeddings. Enç (1996) draws a parallel between *will* and (deontic) modal verbs like *must* or *should*, arguing that the latter also induce a “shift” to the future in their complements as well as in relative clauses:

- (77) a. John must claim that he is sick (Enç, 1996, p. 353, (28)).
 b. John should talk to whoever is guarding the building (Enç, 1996, p. 353, (30)).

Enç concludes that the auxiliary *will* is not a temporal marker but a modal verb, which should be analysed on a par with other modals. Enç proposes that the modals in question behave like Priorean tense operators, always shifting the evaluation time of all their embeddings.

Abusch (1997) also proposes to treat the future auxiliary as a modal, with the consequence that it can be considered an intensional predicate, triggering SOT effects (the same line is taken in (Stechow, 1995)). However, in (Abusch, 1998) she modifies her analysis and proposes a special treatment for the future auxiliary. This is motivated by what Abusch terms a “breakdown in complementarity of present and past tenses” (p. 13), illustrated by the following examples:

- (78) a. On May 21, I will give an automatic A to the first student who turned in a term paper at least 15 pages long (Abusch, 1998, p. 13, (4a)).
 b. On May 21, I will give an automatic A to the first student who turns in a term paper at least 15 pages long (Abusch, 1998, p. 13, (4b)).
 c. Next year, every piece of email which I receive which was received from a friend will be answered (Abusch, 1998, p. 14, (5)).

In these sentences, according to Abusch’s judgements, past and present in the relative clauses embedded under a future matrix verb have the same

temporal interpretation; thus, (78a) and (78b) are equivalent, and in (78c) two simultaneous actions are denoted by past and present. Abusch (1998) proposes a solution in terms of a “now-parameter” – a temporal variable that denotes either the utterance time or a “shifted” reference point for tenses in the scope of an intensional predicate. The future auxiliary is supposed to “expand” the now-parameter infinitely into the future.

The modal analysis of the future auxiliary in English seems to be on the right track; the reader is referred to (Sarkar, 1998) and references cited there for extensive arguments showing that there is at least no evidence against a unified treatment of future and modals. This analysis explains the strange behaviour of present in complement clauses embedded under future, as in (73b), since the same range of interpretations for the embedded tense is available in (77a), where the meaning of the modal verb requires a shift into the future. However, several problems remain unsolved. It is worth noting that any construction with a future meaning results in the same temporal readings for its complement, even when there is no auxiliary *will* involved. Compare the sentences (79a)–(79b) with (79c):

- (79) a. Tomorrow at the party, Mary will say that she is having fun.
 b. Tomorrow at the party, Mary must say that she is having fun.
 c. I promise/am going to say tomorrow at the party that I am having fun.

In addition, dependent temporal readings in relative clauses do not seem to be obtainable as easily as they are in complements. The judgements in (Stowell, 1995) indicate that such readings are only available when the context strongly favours a dependent reading; thus, in Ogihara’s example (74b), our world knowledge tells us that it is relevant for the fish to be alive at the moment of buying, not at the moment of utterance. The situation is the same with modals: judgements do not change if we replace *will* in examples (75) or (74). It seems that only attitude verbs can consistently produce dependent readings for their complements.

In Russian, the situation seems different from English in certain respects. In complements embedded under future, dependent readings for both past and present are also available, which is not surprising for Russian since they are always available under past as well. The relevant example is (80):

- (80) Ivan skazet Maše, čto on bolel / boleet.
 Ivan will-tell Masha-DAT that he ailed / ails
 ‘Ivan will tell Masha that he was/is ill’

However, in relative clauses embedded under future, dependent readings of both past and present seem extremely difficult to get, as shown by the sentences in (81), modelled after Abusch’s examples in (78) and Enç’s (77b):

- (81) a. Priz polučit tot, kto sdelal men'she vsech ošibok.
 Prize will-get that-one who made less all-GEN mistakes
 'The prize will go to whoever made less mistakes (so far)'
- b. Ja postavljaju ocenki tol'ko tem, kto sidit v pervom
 I will-put marks only those-DAT who sits in first
 rjadu.
 row
 'I will only give marks to those (students) who are sitting in the
 front row (right now).'

Judgements on (82) are particularly clear. Imagine that Masha is a little girl. Suppose that her mother, speculating about her daughter's future, utters (82a). Although the extralinguistic context suggests a dependent reading of the embedded tense, this reading is impossible to get, so that Masha's mother can only mean that the person in question loves Masha already. The judgements for (82b), which has a modal instead of future tense in the matrix clause, are the same.

- (82) a. Maša objazatel'no vyjdet замуž za čeloveka, kotoryj eë
 Masha definitely will-get married to person which her
 ljubit.
 loves
 'Masha will definitely marry a person who (now) loves her'
- b. Maša objazatel'no dolžna vyjti замуž za čeloveka, kotoryj
 Masha definitely must get married to person which
 eë ljubit.
 her loves
 'Masha must definitely marry a person who (now) loves her'

Several things might be going on here. First, Russian does not have a future auxiliary of the kind that English does. Instead, Russian uses the present form in perfective aspect or a complex form (see section 1.4.1). The English *will* can be shown to have modal uses in addition to temporal ones; Russian future does not have such properties. Thus, Russian future seems to be different from English. Russian also does not have modal verbs in the English sense. Russian modals do not display a special paradigm, and they can be used in all tenses: present, past, and future. So additional factors might interfere in the comparison.

It is also possible that an inter-speaker variation is present in both languages. It might be that I myself and the Russian speakers that I have consulted are restrictive on dependent readings in future-embedded relative clauses, while Abusch and the speakers she relied on are permissive, in contrast to Stowell. For the moment, I leave the completion of the analysis for future research.

4.4 Summary

In the present chapter, I developed my analysis of SOT in English (and Dutch) and in Russian. The main conclusions that I reached can be summarized as follows:

- In English and other “SOT languages”, a simultaneous reading of past embedded under past is obtained by means of an Agree relation between the T heads of the embedded and the embedding clause; this relation is interpreted as a binding relation. This Agree relation is established when the embedded verb enters the derivation with an unvalued T feature. I follow (Pesetsky & Torrego, 2001, *et seq.*) in claiming that the C head in English bears an instance of the T feature, which normally gets its value from the local T head. When the embedded T is unvalued, an Agree chain is established between the matrix and the embedded T proceeding via the embedded C, and the instances of the T feature on the embedded C and T get their value from the higher T.
- In Russian, the agreement relation in question cannot be established, since in Russian C does not bear an instance of a T feature, so the Agree chain between the two Ts cannot be formed. This is why Russian lacks an “SOT” reading of embedded past of the kind that is found in English. However, in principle a simultaneous reading of past under past is available in Russian. In this chapter, I demonstrate that this reading is different from the English simultaneous past, and argue that in this case the Russian embedded past is interpreted *de re*, along the lines of (Abusch, 1997). As for present under past in Russian, I conclude that the shiftable indexical theory accounts well for its use.

Chapter 5

Long-distance wh-movement in Russian

In this chapter, I turn to a set of problems different from the ones discussed so far, and analyse the contrast between long-distance movement in English and many other languages as opposed to Russian. In section 5.2.1 I present the Russian sentences that illustrate the peculiar situation with wh-movement in Russian: extraction of a wh-phrase out of indicative complement clauses with an overt complementizer is ungrammatical, but the same sentences become grammatical once the complementizer is omitted, and extraction out of subjunctive complement clauses is allowed as well.

In section 5.2.2, I review some of the existing solutions to this problem: (Greenberg, 1988; Meyer, 1997; Bailyn, 1992; Stepanov, 2001; Szczegielniak, 1999). Greenberg (1988) and Meyer (1997) give Subjacency-based accounts of the data; Bailyn's and Stepanov's accounts amount to the suggestions that complement clauses are in fact structural adjuncts in Russian. Szczegielniak (1999) aims at presenting a Phase Theory-based account.

In section 5.3, I develop my own account of long-distance movement in Russian. I propose that the ban on movement out of complement clauses in Russian is due to the specific phase structure of this language: T and not C is the phase head in Russian. I demonstrate that the same structural property of Russian underlies the two seemingly unrelated properties: lack of SOT and extraction difficulties. Section 5.4 is a brief summary.

5.1 Locality of movement

5.1.1 Island constraints

As is well known, long-distance movement of wh-phrases, topicalized elements, and the like is possible but systematically limited in the grammar. An example of licit long-distance movement is given in (1). Here, a wh-phrase is moved out of the deepest embedded complement clause and crosses another complement clause on its way to the matrix clause (the base-generated position of the moved element is denoted by a space). The grammaticality of a construction like (1) is not affected by the number of complement clauses involved (apart from a possible processing overload).

- (1) What did John say that Mary thought that Bill bought _?

In contrast to complement clauses, which allow extraction, certain other environments do not. These were first discussed in the seminal work of (Ross, 1967), who termed them *islands*. Since then, locality phenomena have received considerable attention in linguistic literature, and numerous theories dealing with island constraints were proposed. Islands are usually divided into two groups – strong and weak islands; strong islands are environments that do not allow movement of any element, and weak islands are selective: there are some types of elements that can be extracted from them or produce a lesser degree of ungrammaticality than others. Roughly, arguments can be extracted out of weak islands, while adjuncts cannot: for instance, in (6), the complement *which problem* can be extracted out of the infinitival clause, but not the adjunct *how*. Below I give some examples of strong and weak islands.

1. Strong islands

- (2) * Who did Mary cry after John hit? (Huang, 1982, p. 503) – adjunct island
- (3) *? Who does [a picture of _] hang on the wall? (Stepanov, 2001, p. 18, (1a)) – subject island
- (4) * What the police arrested the man who saw _ was this video. (Szabolcsi & Dikken, 1999, (7)) – complex NP with a relative clause
- (5) * Which man did they consider the rumors that Bob would betray _? (Szabolcsi & Dikken, 1999, (8)) – complex NP with a complement clause

2. Weak islands

- (6) a. ? Which problem did you wonder how PRO to solve _?
 b. * How did you wonder which problem PRO to solve _?
 (Huang, 1982) – wh-islands
- (7) a. How do you think that I should cook this stuff _?
 b. * How don't you think that I should cook this stuff _?
 c. * How do you think that I shouldn't cook this stuff _?
 (Starke, 2001, p. 5, (6)) – negation islands

5.1.2 Theories of locality

Within the Principles and Parameters framework, locality of movement was accounted for by the theory of Subjacency (Chomsky, 1973, 1977), which later evolved into the theory of Barriers (Chomsky, 1986). The common intuition behind these theories is that a phrase moved out of its local domain can only cross a limited number and/or a specific combination of certain nodes. If this condition is violated, i.e. if the phrase moves out “too far” (in structural terms), the sentence becomes ungrammatical.

In Subjacency, the “offending” nodes were called *bounding nodes*, and were subject to parametric variation: for English, these were NP and S (IP), for Italian NP and S' (CP) (Rizzi, 1982). The rule stated that movement could not cross more than one of these at once. Constructions like (1) were salvaged by the idea that movement is successive-cyclic: the moved element does not get from its base position into the matrix clause in one fell swoop, but proceeds in cycles, each CP being a cycle for the application of movement. Thus, the moved phrase uses each intermediate Spec,CP as an intermediate landing site, an “escape hatch”. Only one IP at a time is crossed by movement of *what* in (1); in contrast, extraction of a complex NP like (4) and (5) has to cross both an NP and an IP before reaching the matrix Spec,CP.

Subjacency also provided an explanation for wh-islands: in a sentence like (6b), the moved wh-phrase cannot use the Spec of the embedded C as an escape hatch, because this position is already occupied by another wh-element. Thus, it is forced to move directly to the matrix position, crossing more than one IP on its way. However, this approach does not offer an explanation for the contrast in grammaticality between sentences like (6a) and (6b), where different types of elements are extracted across another wh-phrase.

A different strategy of accounting for locality of movement was provided by the Empty Category Principle (the ECP), first proposed in (Chomsky, 1981) and developed in (Lasnik & Saito, 1984, a.o.). This principle relies on the idea that movement always leaves behind a trace, that is, an empty category. The ECP defines the conditions under which empty categories are licit; thus, it does not deal with movement directly, but rather with positions that traces of movement can occupy. The general requirement on

empty categories is that they should be properly governed, that is, either θ -governed (lexically governed), or governed by its antecedent – an element it is coindexed with and c-commanded by (in case of movement, the moved element itself) (see (Chomsky, 1986; Aoun & Sportiche, 1981) for the exact definitions). Traces of complements satisfy the first condition: they are always in the government domain of the verb. By contrast, subjects and adjuncts are not lexically governed, hence they have to be governed by an antecedent. Now variation in acceptability of *wh*-island extraction could be accounted for: an adjunct trace, like that in (6b), is not governed by its antecedent.

A similar intuition lay behind the Condition on Extraction Domains (the CED), which was proposed in (Huang, 1982) to account for the impossibility of extraction out of subject and adjunct clauses. The CED states that no element can be extracted from a domain that is not properly governed. This means that complements, which are lexically governed by the verb, are transparent for extraction, but subjects and adjuncts, for lack of a governor, are opaque.

The Barriers approach (Chomsky, 1986) aimed at providing unified locality conditions for movement, ECP and Case assignment. The idea behind Barriers is that no structural elements are obstacles to movement or government (*barriers*) by definition. Rather, certain maximal projections become barriers due to the syntactic configuration in which they are found. In defining these configurations, Chomsky (1986) relies on the notion L-marking, which he identifies as “direct θ -marking by a lexical category” (Chomsky, 1986, p. 11). This relation distinguishes between elements receiving a theta-role and those that do not; in effect, between complements of lexical items and all other elements. A maximal projection can be an inherent barrier if it is not L-marked – with the exception of IP which does not acquire barrierhood in this case. Alternatively, it can become a barrier by inheritance if it immediately dominates an inherent barrier. Similarly to Subjacency, movement cannot cross more than two barriers at a time, and government cannot cross more than one barrier. Thus, for an empty category to be antecedent governed, no barriers can intervene between the category and its antecedent. In addition, government by the antecedent will be blocked if another potential governor – a *minimality barrier* – intervenes.

The theory of Relativized Minimality (Rizzi, 1990) also builds on the concept of intervention and elaborates on the ECP; however, Rizzi (1990) “relativizes the notion of “minimality barrier” in the sense that a category intervening across a non-rigid barrier can block antecedent government but *only* if it is of the same type as the antecedent whose government is being blocked” (Hornstein & Weinberg, 1995, p. 279). Types of interveners involved are heads versus specifiers and, within the latter type, A- and A'-positions: a potential head-governor blocked a head government relation, etc. Relativized Minimality thus can account for a broad array of weak

islands. For instance, in negation islands, the negation *not* can act as an intervener for antecedent government of an A' trace if it is analysed as the Spec of a negation projection, thus in an A' position.

The minimalist program takes a different stand with respect to the nature of movement, as compared to Government and Binding theory: movement is not free but triggered by features that need to be checked. For instance, wh-movement is triggered by an unchecked wh-feature on C, which attracts the wh-word carrying a matching instance of the feature. In this framework, relativized minimality effects were restated in terms of the Minimal Link Condition (MLC), which states that “K attracts α only if there is no β , β closer to K than α , such that K attracts β ” (Chomsky, 1995, p. 311, (110)).

Rizzi (2002) elaborates his theory of Relativized Minimality in similar terms, dispensing with the notion of antecedent government. Instead, he proposes that “local relations must be satisfied in a minimal configuration, the smallest configuration in which they can be satisfied” (Rizzi, 2002, p. 224). This idea resembles the MLC of (Chomsky, 1995). However, a violation of the MLC requires that the intervener and the moved element have the same featural makeup. Rizzi (2002) does not appeal to the feature makeup of the intervening element, arguing that this approach is too selective. While the MLC can account for wh-island effects, other types of weak islands, such as islands induced by negation and quantificational adverbs, are left out. On the other hand, Rizzi demonstrates that operating with only the three types of interveners defined in Relativized Minimality (heads, A- and A'-specifiers) is not selective enough; he proposes a more elaborate typology of specifiers based on different classes of features they license, such as argumental (person, number), quantificational (Wh, Neg), etcetera.¹ Besides, Rizzi (2002) argues that there are further factors involved in extractability of elements from weak islands. He notes that (Cinque, 1990) “observed that not all arguments are extractable from Weak Islands: in order to be successfully extracted, arguments must have special interpretive properties, they must be specific, or presupposed, or D(discourse)-linked, in the sense of Pesetsky (1987), with the range of the variable pre-established in previous discourse (...)” (Rizzi, 2002, p. 230)).

The minimalist program does not readily supply an explanation for strong island effects, such as subject and adjunct islands. The CED, which was designed to deal with these issues, was rejected in minimalism since the relation of government, upon which it relied, was not part of the system any more. Several theories have been proposed to solve the problem. Thus, the approach of (Nunes & Uriagereka, 2000) builds on the idea of multiple Spell-out developed in (Uriagereka, 1999). Multiple Spell-out is claimed to be a consequence of the Linear Correspondence Axiom (LCA) proposed in (Kayne, 1994), which derives linear precedence from c-command rela-

¹A further elaboration of this approach is found in (Starke, 2001), who proposes a unified theory of strong and weak islands based on the idea of feature intervention.

tions: “A lexical item α precedes a lexical item β iff α asymmetrically c-commands β ” (Nunes & Uriagereka, 2000, p. 23). (Nunes & Uriagereka, 2000) point out that due to the way Merge proceeds, some parts of the structure – in particular, subjects and adjuncts – “must be assembled in a separate derivational space before being connected to the rest” (Nunes & Uriagereka, 2000, p. 22). But the resulting structure is then impossible to linearize, since the system cannot establish the c-command relations between the items inside the “separately assembled” elements and the rest of the structure.

The solution proposed in (Uriagereka, 1999) is that these elements are linearized independently and then sent to Spell-out prior to being integrated in the structure. In other words, multiple applications of Spell-out are required, and subjects and adjuncts have to be sent to Spell-out in the process of derivation, while the structure is still incomplete. (Nunes & Uriagereka, 2000) propose that, as a consequence, the internal structure of subjects and adjuncts becomes invisible to syntactic operations, and no element can be moved out of them.

5.1.3 Locality of movement and successive cyclicity in Phase Theory

In section 2.1.2, I outlined the main principles of Phase Theory of Chomsky (2005, 2006). In this section, I return to the account of successive cyclicity and locality that is provided by Phase Theory.

According to (Chomsky, 2005), Phases – CP and v*P, possibly also DP – are “chunks” of syntactic structure that get sent off to the interfaces in the process of the sentence computation; after that, only the phase edge, that is, the phase head and its Spec, remain visible for subsequent operations. Chomsky argues that all syntactic processes, including various instances of movement, are supposed to apply at phase edge. Consequently, long-distance movement has to be successive-cyclic: no element is visible if it is inside completed a phase, so it has to move out to the phase edge in order to be moved further.

Chomsky (2005) outlines an account for adjunct and subject islands, although he does not go into much detail on these issues. Island effects are explained in terms of search domains for probe-goal relations. Chomsky claims that “[f]or minimal computation, the probe should search the smallest domain to find the goal: its c-command domain” (p. 12). However, in XP-YP constructions where neither XP nor YP is a head the minimal search domain is not uniquely defined, and “the “wrong choice” yields island effects”. However, Chomsky (2005) does not propose a safe method to identify the right choice; he mentions Kayne’s connectedness principle and government, but argues that both are not optimal since they involve additional stipulations.

Adjunct island effects are argued to follow from the treatment of ad-

juncts in (Chomsky, 2004); in this account, adjuncts are taken to be “entered into the derivation by pair-Merge instead of set-Merge to capture the fundamental asymmetry of adjunction, then simplified to set-Merge at the point of Transfer, thus permitting phonetic linearization and yielding “late-insertion” effects at the semantic interface” (Chomsky, 2005, p. 12). As for subject islands, Chomsky argues that not all subjects induce island effects, but only subjects of active verbs, that is, those which have a v^* phase.

As already noted in section 2.1.2, Chomsky (2005) dispenses with wh -features, or indeed any features that are usually supposed to trigger A' -movement (like focus features). He argues that movement to A' -positions at phase edges is unrestricted and driven by edge features (EFs) of the phase head; any phrase can be moved to a phase edge position, and whether or not the resulting structure converges depends on the possibility of interpreting the resulting configuration. The procedure for wh -movement is outlined as follows: “(. . .) suppose that the moved phrase is labeled by an interpretable interrogative wh -feature. Then it will have to reach the right position in the left periphery for interpretation, or be associated with such a position by some other operation” (Chomsky, 2005, p. 17).

Importantly, this approach allows Chomsky (2005) to get rid of an important problem for the theory of successive-cyclic movement, namely the problem of intermediate movement positions. This problem arises from the minimalist assumption that all movement is feature-driven: if this is the case, then what are the features that trigger movement of the extracted element to the intermediate “landing sites”, or “escape hatches” (which in Phase Theory are phase edges)? The problem is even more grave than that, since, as Rizzi (2004) remarks, “[t]he paradox of these intermediate positions is that, on the one hand, they must autonomously cause a movement step (if we want to take seriously the idea that each step is locally determined, with no “look-ahead” to subsequent derivational steps), and at the same time we should make sure that movement will not stop there” (Rizzi, 2004, p. 9). Not having any attracting features involved in A' -movement means that the problem simply does not arise.

However, Chomsky’s hypothesis seems to raise problems, as discussed e.g. in (Slioussar, 2007), who lists several issues this approach leaves open:

- If the wh -phrase is endowed with an interpretable wh -feature, it is not obvious why it is necessary for it to move in order to be interpreted. If indeed it has to move, treatment of the wh -phrases that remain in situ (as some wh -phrases in English and all wh -phrases in some other languages do) becomes a problem. Chomsky (2005) remarks that the wh -phrase can be not only moved to a position in the left periphery, but also be connected to it “by some other operation”. However, Chomsky does not specify what kind of operation this could be, though he does refer to some previous work for possible analyses of wh -in-situ (Chomsky, 2005, p. 17, fn. 46).

- It is not specified which is the position that the *wh*-phrase has to end up in. As Slioussar (2007) notes, “[i]f there is no feature-matching in *wh*-features and *wh*-movement is purely EF-driven, it will be difficult to connect a *wh*-phrase to a particular head in the C-domain (or to a particular feature of C itself). Thus, an additional interface rule will have to be introduced” (p. 34).

Phase-based approaches to long-distance movement that do rely on *wh*-features were put forward in (McCloskey, 2002) and in (Rizzi, 2004), who deal specifically with the problem of intermediate movement positions. McCloskey (2002) provides evidence that points towards feature-driven movement to intermediate Spec,CP’s. He discusses data from Irish, where the complementizer of an embedded clause changes depending on whether or not successive-cyclic movement has taken place – and, more generally, “whether or not a relation of A’-binding reaches into the clause” (McCloskey, 2002, p. 188-189). It is also important what type of A’-binding that is: if it is binding of a trace (in other terms, a copy of a moved element), the complementizer has the form conventionally represented as *aL* (example (8a)); if it is binding of a resumptive pronoun, the complementizer has the form represented as *aN* (example (8b)). Otherwise, the complementizer *go* is used, as in (8c).

- (8) a. Creidim gu-r inis sé bréag.
I-believe go-[PAST] tell he lie
‘I believe that he told a lie.’ (McCloskey, 2002, p. 185, (3))
- b. an t-ainm a hinnseadh dúinn a bhí _ ar an áit
the name *aL* was-told to-us *aL* was on the place
‘the name that we were told was on the place’ (McCloskey, 2002, p. 185, (5))
- c. an ghirseach a-r ghoid na síogaí í
the girl *aN*-[PAST] stole the fairies her
‘the girl that the fairies stole away’ (McCloskey, 2002, p. 189, (9b))

Demonstrating a wide range of Irish data, McCloskey (2002) argues against a proposal expressed in (Heck & Müller, 2000) that intermediate movement steps are not feature-driven but result from an Optimality-theoretical constraint ranking that governs syntactic derivations.²He proposes that the intermediate C heads supporting long-distance relations (which have the form *aL* or *aN*) bear an EPP feature (which would correspond to Edge Features of Chomsky’s later work). In addition, if long-distance

²(Heck & Müller, 2000) argue that the constraint on the application of movement known as Last Resort is overridden at intermediate stages of the derivation by a higher-ranked constraint which they call “Phase Balance”, which requires that at the completion of every phase all unchecked features be accessible for subsequent checking.

movement occurs, the intermediate C also bears a feature *Op* (an operator feature), which McCloskey calls a kind of a “pseudo” wh-feature, in the sense that it is not interpretable; this feature agrees with a matching feature on the wh-phrase. In that case, C is spelled out as *aL*. If C does not bear an *Op* feature, then its EPP feature is satisfied by merging a (phonologically empty) operator binding a resumptive pronoun, which is due to economy since Move is more costly than Merge. Then C is spelled out as *aN*. So, according to McCloskey’s analysis, there is a difference in the form of the complementizer depending on whether or not it bears an operator feature that attracts the moving wh-phrase.

Rizzi (2004) provides a general discussion on the form of chains resulting from (long-distance) A’-movement. His theory is based on a distinction between s-selection positions (positions in which elements are semantically selected, such as thematic positions for arguments) and criterial positions (positions in which scope-discourse properties are expressed). An A’-chain is always headed by a criterial position, while its lowest position must be an s-selection position. Between them, there can be a series of intermediate positions, the “landing sites” of long-distance movement. Rizzi (2004) provides several arguments that point to the existence of these positions, such as the complementizer agreement data discussed by McCloskey (2002) for Irish and attested in many other languages, or reconstruction effects demonstrated in (Lebeaux, 1988): in (9a) not only *John* but also *Bill* can be the antecedent of the anaphor *himself*, which is explained if a copy of the DP containing the anaphor is available at the embedded Spec,CP.

- (9) a. [Which picture of himself_{i, k}] does [John_k think [that [Bill_i likes ___]]]?
 b. [Which picture of himself] does [John think [<which picture of himself> [that [Bill likes <which picture of himself>]]]] (Rizzi, 2004, p. 8, (25d-e))

Rizzi (2004) proposes that all movement is feature-driven, hence movement to the intermediate positions is feature-driven as well. Movement to the final (criterial) positions is driven by criterial features, such as wh-features, which have purely formal counterparts located at the intermediate movement positions (like the “pseudo” wh-features of (McCloskey, 2002)). These formal features “have the role of bringing the moved phrase closer to the target, in accordance with locality principles”; Rizzi concludes that “[n]atural language syntax includes the latter featural option in order to implement intermediate movement, so that a potentially distant criterial position can be reached in successive steps, each of which satisfies locality requirements” (Rizzi, 2004, p. 10).

A very similar view is expressed in (Pesetsky & Torrego, 2004b) (see section 2.1.3) who argue that the target C head of the wh-movement bears an interpretable, unvalued instance of a wh-feature (which they term “Q

feature”), while the intermediate Cs bear uninterpretable, unvalued instances of the same feature (the Q feature on the wh-phrase itself is uninterpretable but valued). This ensures that the wh-phrase is attracted into an intermediate position by an unvalued feature but cannot stop there since the attracting feature is still uninterpretable. In this thesis, I opt for an approach to wh-movement and intermediate movement positions along these lines.

5.2 Locality of movement in Russian

The properties of long-distance movement in Russian have been a puzzle for linguists for some time. Since (Ross, 1967; Comrie, 1971), it has been noted that long-distance movement in Russian is very restricted: in addition to the island constraints established for English and many other languages (adjunct island, complex NP island, etc), Russian disallows extraction out of indicative complement clauses. This is unexpected since complement clauses are normally transparent for long-distance movement in English and languages like it. None of the existing theories of locality yield a ban on movement in this configuration. In this section, I will present the relevant data and give an overview of the existing accounts.

5.2.1 Russian data: the puzzle

The English sentence (10) illustrates long-distance extraction of a wh-element out of a complement clause:

(10) Who do you think that John invited _?

If we try to translate (10) into Russian, we run into difficulties. The literal translation would have to be (11):

(11) ?* *Kogo ty думаеш, что Ivan пригласил _?*
 Whom you think that Ivan invited
 ‘Who do you think that Ivan invited?’

Here the wh-word *kogo* ‘who’ (accusative) is fronted from its position in the complement clause introduced by the complementizer *что* ‘that’, just as in its English counterpart. However, this sentence is judged by most speakers to be degraded.³

³It is worth noting that this is not a property of the complementizer *что* as a lexical item. For instance, the complementizer *будто* (synonymous to *что* but conveying a larger degree of the speaker’s uncertainty about the content of the complement) is also opaque for extraction:

- (i) * *Vot devuška, kotoruju ty skazal, budto on ljubit.*
 Here girl whom you said that he loves
 ‘Here is the girl whom you said that he loves’ (Comrie, 1971, p. 344, (34))

Interestingly, the same construction improves if the complementizer is not overtly present in the structure:

- (12) Kogo, ty dumaeš, Ivan priglasil _?
 Who you think Ivan invited
 ‘Who do you think Ivan invited?’

In addition, extraction out of subjunctive complement clauses is also grammatical:

- (13) Kogo ty xočeš, čtoby Ivan priglasil?
 Who you want that-SUBJ Ivan would-invite
 ‘Who do you want Ivan to invite?’

Here, the sentence is introduced by the complementizer *čtoby* that consists of the indicative complementizer *čto* and the particle *by* which is part of the subjunctive form of the verb (see section 1.4.1). Surprisingly, extraction across *čtoby* is allowed, in contrast to *čto*.⁴

It should be noted that the judgements on the data discussed above are not uniform. While some speakers reject examples like (11) as completely ungrammatical, others accept them in colloquial speech. This interspeaker variation is also reflected in the literature. While most authors claim that examples like (11) are ungrammatical, a few consider them possible. Yadroff (1991, 1992) claims that in colloquial Russian, which should be distinguished from the highly prescriptive “codified literary Russian”, long-distance extraction of any kind is widespread. Van Gelderen (2001) reports disagreement among speakers regarding the grammaticality of long-distance *wh*-movement constructions; according to her, these

⁴Bailyn (2003, 2006) claims that the *that-t* effect holds for the Russian *čtoby*-clauses, giving examples like (i) (Bailyn, 2003, p. 11, (28b)):

- (i) *Kto ty xočeš, [čtoby t_i v]ljubilsja v Iru?
 who-Nom you want that fall in love (to) Ira
 ‘Who do you want to fall in love with Ira?’

Some speakers do find sentences like (i) worse than their counterparts with object extraction. However, the judgements are not as categorical as Bailyn suggests. For other speakers, extraction of the subject is grammatical; for instance, an internet search for relevant constructions produced a number of sentences like (ii) (from the memoirs of A.D. Sakharov, available at http://www.sakharov-archive.ru/Raboty/V_Glava_2_30.htm):

- (ii) A kto, ty xočeš, čtoby vzjal tebja zamuž?
 And who you want that-SUBJ take you married
 ‘So who do you want to get married to?’ (lit. Who do you want to take you as a wife?)

I conclude that there is not enough evidence to postulate the existence of *that*-trace effect in Russian *čtoby*-clauses.

constructions are never degraded for some speakers while they display various degrees of degradation for others.

After conducting several speaker questionnaires, and following my own intuitions, I conclude that the constructions in question are indeed degraded, but not as severely as strong island violations, e.g. adjunct island violations:

- (14) *Kogo Maša zaplakala, kogda Petja udaril?
 Whom Masha cried after Petya hit
 ‘*Who did Masha cry after Peter hit?’

The question of course arises whether Russian has any other way of expressing the intended meaning of (11). In fact, the fully grammatical counterpart of a long-distance wh-question in Russian is a construction similar to the German wh-scope marking (“was-w”) construction, which uses *kak* “how” as the wh-scope marker:

- (15) Kak ty dumaeš, kogo Ivan priglasil?
 How you think whom Ivan invited
 ‘Who do you think Ivan invited?’

As mentioned above, Russian displays traditional island constraints, such as adjunct island, complex NP island, etc. Stepanov (2001) makes the claim that Russian does not respect the subject condition, i.e., extraction out of a subject clause is grammatical in Russian. However, this claim does not seem tenable. Stepanov (2001) provides the following examples (judgements his):

- (16) a. S kem by ty xotel čtoby govorit’ bylo by
 With whom SUBJ you want that-SUBJ to-speak be SUBJ
 odno udovol’stvie?
 one pleasure
 lit. ‘With whom would you want that [to speak _] were sheer pleasure’
 b. Čto by ty xotel čtoby kupit’ ne sostavljalo by
 what SUBJ you want that-SUBJ to-buy not constitute SUBJ
 nikakogo truda?
 no labor
 lit. ‘What would you want that [to buy _] would not be any trouble?’

Speakers of Russian I have consulted (as well as myself) do not accept these sentences as absolutely grammatical: (16a) is degraded, though it might be possible marginally, and (16b) is quite bad. Besides, the examples raise additional questions: first, (16a) involves the subject of the verb *byt’* “to be”, so it is not the subject of a transitive verb (cf. (Chomsky, 2005) who

claims that only subjects of transitive verbs induce subject island effects in English). The verb in the *čtoby*-clause in (16b) is formally transitive but the expression *ne sostavl'jat' nikakogo truda* “be no trouble” is an idiom; in any case, the sentence is worse than the first.

Furthermore, the discussion of subject clauses in Russian is hampered by the fact that it is very difficult to compose reliable examples with extraction out of a subject. First, hardly any transitive verbs can easily take a clausal subject, either finite or infinitival (I have not been able to construct any good examples; see also section 2.1.3 for a discussion of subject clauses in Slavic). In addition, examples of this kind are generally considered unreliable because of the doubtful status of subject clauses (see Koster, 1978). Most often, picture NP subjects are used to demonstrate (the impossibility of) extraction in English, as in (3) on page 126, repeated here as (17).

- (17) *? Who does [a picture of _] hang on the wall? (Stepanov, 2001, p. 18, (1a))

But in Russian, extraction of complements out of NPs is generally quite bad, even when the NP is in an object position:

- (18) ?* Kogo ty videl fotografiju?
 whom you saw photograph
 ‘Who did you see a photo of?’

Thus, for independent reasons it is hard to test subject islands in Russian.

5.2.2 Existing accounts

There have been several attempts at explaining the Russian extraction data. I will start from older, pre-minimalist accounts and then turn to more recent work, paying special attention to the account by Stepanov (2001), which offers the most detailed overview of the data and a fully-fledged theory of Russian complement clauses. I will also mention Szczegielniak’s (1999) phase theory-based account.

Ross (1967), Comrie (1971, 1973)

The first mention of the Russian extraction facts is found as early as (Ross, 1967). Ross argues that no syntactic processes can reach into Russian *čto*-clauses; this is true for rules such as reflexivization (the antecedent of the pronoun *sebjā* cannot be located outside the clause) and genitive of negation introduction, and also for relativization, as shown in (19) (Ross, 1967, p. 464, (6.222)): no element can be moved (or “chopped”, in Ross’s terminology) from a *čto*-clause.

- (19) * Vot ženščina, kotoruju ja znal, čto on ljubil.
 This woman which I knew that he loved

'Here is the woman who I knew that he loved'

The topic is further developed by Comrie (1971), who makes the important observation that *wh*-extraction out of *čtoby*-complements is grammatical:

- (20) a. *Vot kniga, kotoruju ja prikazal, čtoby on pročital.*
 This book which I ordered that-SUBJ he read
 'Here is the book that I ordered him to read'
- b. *Čto vy xotite, čtoby ja prinēs?*
 what you want that-SUBJ I bring
 'What do you want me to bring?' (Comrie, 1971, p. 346-347, (46)-(47))

Comrie proposes a distinction between finite complements, which are impenetrable for transformations, and infinitival clauses, which are transparent. He demonstrates that *čtoby*-clauses behave like finite complements in many respects except for extraction, and concludes that they should be treated as exceptions. Similarly, (Comrie, 1973) examines various types of extraction out of different types of clauses in Russian in order to provide an explanation in terms of Extended Standard Theory. His conclusion is that some movement transformations cannot cross the boundaries of embedded sentences with finite verbs or infinitive constructions with a complementizer.

Subjacency-based accounts: Greenberg (1988), Meyer (1997)

Greenberg (1988) examines relativization out of *čto*- and *čtoby*-clauses in Russian and offers an account of their grammaticality based on Chomsky's Barriers approach. He proposes that in sentences like (21), the complementizer *čto* does not allow the movement of the relative pronoun into Spec,CP. Therefore, the relative pronoun has to cross two barriers – IP, which is the most deeply embedded tensed IP in the sentence and therefore an inherent barrier, and CP, which is a barrier by inheritance. Therefore, (21) is ungrammatical⁵:

- (21) * *ogurcy, kotorye ja obeščal, čto prinesu.*
 cucumbers which I promised that I-will-bring
 'the cucumbers which I promised to bring' (Greenberg, 1988, p. 189, (10))

In case of extraction across the complementizer *čtoby*, the deepest embedded IP is subjunctive, so Greenberg considers it to be a kind of intermediate case between a tensed and an infinitival IP. This is why it does not induce strong ungrammaticality.

⁵In Greenberg's and Meyer's examples, I simplify the glosses for the sake of exposition.

Another Barriers-based explanation of Russian extraction facts is offered by Meyer (1997). Meyer adopts Erteschik-Shir's (1973) theory of "bridge verbs" which classifies matrix verbs introducing complement clauses into those that allow extraction easily ("bridge verbs") and those that do not. Verbs with greater "semantic complexity" are usually worse bridges.

Discussing the Russian data, Meyer notes that extraction out of *čtoby*-complement clauses is only acceptable with volitional verbs (and other verbs, e.g. verbs of speech, with a volitional meaning):

- (22) a. Čem Saša xotel, čtoby deti zanimalis'?
with-what Sasha wanted that-SUBJ children be-occupied
'What did Sasha want the children to do?'
- b. Vot kniga, kotoruju nužno/nel'zja, čtoby oni
Here book which necessary/not-allowed that-SUBJ they
čitali.
read
'This is the book which they should read / must not read' (Meyer, 1997, p. 167, (7a-b))

However, subjunctive can be used in Russian complement clauses without a volitional verb. In that case, subjunctive conveys "a meaning of doubt towards the content of the complement clause" (p. 167), and it is possible when the matrix clause contains negation or a rhetorical question, or with matrix verbs like *somnevat'sja* "doubt" or *otricat'* "deny". In such cases, indicative is also possible:

- (23) Ja ne dumaju, čtoby/čto on takoe soveršil.
I not think that-SUBJ/that he such commit
'I don't think he could commit something like that' / 'I don't think he could have committed something like that'. (Meyer, 1997, p. 168, (10))

Meyer argues that in this case, the complement clause does not become more transparent for movement than a normal *čto*-clause:

- (24) Kogo načal'nik usomnilsja, ??čto/*čtoby sekretar' vyšvyrnul?
Whom boss doubted that/that-SUBJ secretary kick out

'Who did the boss doubt that the secretary kicked out/might kick-out?' (Meyer, 1997, p. 168, (11))

However, in cases where a matrix predicate can acquire a volitional meaning and introduce a subjunctive complement clause, extraction becomes easier with *čtoby* when compared to *čto*:

- (25) Kogo vam važno, čtoby/*čto vse škol'niki
 Whom to-you important that-SUBJ/that all schoolchildren
 uvažali?
 respect
 'Whom is it important to you that all schoolchildren admire/admired?' (Meyer, 1997, p. 169, (15))

Meyer proposes that the crucial property of the volitional predicates is that the action of the embedded clause is necessarily posterior to the action of the matrix clause. In that case, there must be a difference in the possibility of extraction from *čto*-clauses as well, depending on whether the action expressed by the embedded predicate is posterior or simultaneous/preceding the matrix predicate time. He claims that extraction in case of posteriority is indeed slightly easier, although the speakers he consults do not provide a uniform picture, and in any case, these examples are not nearly as good as *čtoby*-extraction.

- (26) [Ne znaju,] kuda on ?poobeščal/??dumaet, čto my poedem.
 [Not know] where he promised/thinks that we will-go
 'I don't know) where he promised/thinks that we will go' (Meyer, 1997, p. 171, (27))

To explain these facts, Meyer refers to Brecht (1977) who gives arguments in favour of the view that the complementizer *čtoby* is in fact not a separate complementizer but consists of the complementizer *čto* plus the subjunctive particle *by*. Furthermore, Meyer (1997) claims that in colloquial Russian the two can sometimes be separated (Zemskaja, Kitajgorodskaja, & Shirjaev, 1981):

- (27) Glavnoe čto reka by rjadom byla.
 Main-thing that river SUBJ nearby be
 'The most important thing is that there should be a river nearby' (Meyer, 1997, p. 173, (31a))

Meyer argues that if the conjunction is "split", extraction becomes problematic:

- (28) a. *Kogo on napisal, čto v Novyj God by ne zabyli?
 Whom he wrote that in New Year SUBJ not they-forget
 'Who did he write that they shouldn't forget at New Year?'
 b. ?Kogo on napisal, čtoby v Novyj God ne zabyli?
 Whom he wrote that-SUBJ in New Year not they-forget
 (Meyer, 1997, p. 173, (32))

He concludes that "overt movement of *by* from I^0 to C^0 opens up a barrier, namely IP" (p. 173). But with volitional predicates, this is required by the

lexical properties of the verb (the verb lexically selects for subjunctive). In other cases (e.g. doubt predicates) this does not happen. In case of volitional predicates, a tense feature is transported from the verb to the complementizer. In other cases no feature sharing takes place. Thus, with “real” volitional predicates neither IP nor CP is a barrier. With verbs of speech functioning as volitional predicates, however, CP can be a barrier, depending on the “bridge status” of the verb. In case of an indicative complement, IP will always be a barrier, and CP is again subject to “bridge-ness” variation, as well as to inter-speaker variation.

Bailyn (1992)

Bailyn (1992) proposes that “the doubly filled Comp filter prevents *wh*-movement thorough an embedded Spec C position if the C⁰ position is filled at LF” (p. 320-321, fn. 10). The difference between English and Russian is that in English, the complementizer *that* can be dropped in syntax, and by assumption is always dropped at LF. In Russian, on the other hand, the complementizer cannot be dropped at LF if it is present in syntax. For this reason, movement only obtains when *čto* is dropped in syntax.

Bailyn (1992) proposes that this is due to the fact that Russian indicative complement clauses have to move out of their base-generated position to escape case assignment by the verb, due to the Case Resistance Principle (Stowell, 1981), unless *čto* is dropped. After the movement occurs, they become structural adjuncts, and *wh*-movement out of them is therefore illicit because it would violate the CED. *Čtoby*-clauses do not have to undergo movement, because the complementizer *čtoby* is deletable at LF.

Stepanov & Georgopoulos (1997)

The analysis by (Stepanov & Georgopoulos, 1997) is based on the contrast between extraction out of indicative and subjunctive embedded clauses. Stepanov & Georgopoulos propose that a feature $\pm R$ (ealis) is assigned to all phrase markers. Each phrase marker must have a uniform R-status, which means that Merge can only combine elements with an equal R-status (+R or -R). Instances of irrealis (-R setting) are questions and subjunctive mood. Another theoretical assumption Stepanov & Georgopoulos make is that in the case of long-distance extraction Russian, in contrast to e.g. English, requires Move (fronting the *wh*-phrase) to apply before Merge (of the embedded clause and the matrix clause).

A long-distance *wh*-question is claimed to have two variants of derivation. In a sentence like (29) the clause *ty dumaeš čto* (which is supposed to have a null IP, similar to the structure proposed for sluiced constituents) is adjoined to the IP node of *Kogo Elena privedët*. This derivation must crash because of the conflict between the +R setting of the complementizer *čto* and the -R setting of the embedded clause involving *wh*-movement [*Kogo*

[*Elena privedët t*]]. This conflict is resolved if the (overt) complementizer is eliminated. In contrast to *čto*-clauses, extraction out of a subjunctive clause does not cause a conflict because the subjunctive complementizer *čtoby* is marked -R.

- (29) * Kogo ty dumaeš, čto privedët Elena?
 Who you think that will-bring Elena?
 ‘Who do you think that Elena will bring?’

For *čto*-clauses, the conflict can be avoided if another derivational path is chosen, namely adjoining the first clause to the CP and not to the IP of the second clause, because in that case neither of the complementizers is in the scope (c-command domain) of the other one. This derivation is used to yield the partial *wh*-movement construction, marginally acceptable in the authors’ view:

- (30) ? Ty dumaeš, čto kogo Elena privedët?
 You think that who Elena will-bring
 ‘Who do you think that Elena will bring?’

Stepanov (2001)

In (Stepanov, 2001), a different analysis of the same phenomena is given. In contrast to (Stepanov and Georgopoulos 1997), here Stepanov is concerned only with sentences where the complementizer *čto* is present overtly; he argues that although without the complementizer the sentences become acceptable, this is due to the fact that the matrix clause here is in reality a parenthetical constituent inserted into a root question.

(Stepanov, 2001) starts by exploring constructions where the embedded clause is (optionally) preceded by a pronominal correlative *to* ‘it’, ‘that’:

- (31) Pëtr utverždal (to), čto Ivan ljubit Mašu.
 Petr claimed it that Ivan loves Masha
 ‘Petr claimed that Ivan loved Masha’ (Stepanov, 2001, p. 143, (16))

However, not just any verb can appear in this construction:

- (32) * Pëtr sčitaj to, čto Ivan ljubit Mašu.
 Petr believed it that Ivan loves Masha
 ‘Petr believed that Ivan loved Masha’ (Stepanov, 2001, p. 145, (21))

Stepanov lists two ‘natural classes of predicates’, one of which includes verbs that can (but do not have to) appear with *to*+clause, and the other those that cannot. He then proposes that a correlative pronoun is always present in the structure, even when it is not phonologically realised. He

follows (Giejgo, 1981) who suggests for Polish that the correlative is always present in finite complements, but can delete at PF. Stepanov makes a different claim: there are two correlatives in the system, overt *to* and silent TO. The silent TO necessarily occurs with verbs of the second class (*dumat'* “think”, *polagat'* “suppose” etc), but can also occur with verbs of the first class, which accounts for the cases where there is no overt correlative with a first class verb. Thus, TO appears in the following sentences:

- (33) a. Pětr utverždal TO/to, čto Ivan ljubit Mašu.
 b. Pětr sčital TO, čto Ivan ljubit Mašu.

The crucial difference between *to* and TO is that *to* always bears the Case feature, while TO can be drawn from the lexicon with or without the Case feature. Verbs from the first class are structural Case assigners/checkers and take a direct object; appearing with them, TO bears Case. The embedded clauses are adjoined to these correlatives, which makes them impenetrable for extraction (thus, they are structural adjuncts). This structure is termed “NP-shell” – a notion Stepanov borrows from Müller (1995).

In order to account for the possibility of extraction out of embedded subjunctive clauses, Stepanov claims that volition predicates, which introduce subjunctive clauses with the complementizer *čtoby*, do not select NP-shells. This is demonstrated by the fact that these predicates never occur with the correlative *to*, even though, according to Stepanov, some of them are structural Case assigners:

- (34) ?* On xočet to, čtoby Ivan vyigral
 He wants it that-SUBJ Ivan win
 ‘He wants Ivan to win’ (Stepanov, 2001, p. 152, (34))

After providing the explanation to the complement clause extraction phenomena, Stepanov proceeds to analyse wh-scope marking in Russian. He claims that this construction has essentially the same structural configuration as the NP-shells. According to him, the wh-scope marker *kak* “how” is a head that takes the finite clause as a complement (thus differing from NP-shells which involve adjunction of the clause). The wh-scope marker is Caseless and so it is selected by the same type of verbs that silent TO. Later in the derivation, the wh-scope marker moves to the matrix C.

Szczegielniak (1999)

Szczegielniak (1999)⁶ proposes a theory of that-t effects in different languages, including English, German, Dutch, Polish, and Russian. His proposal is based on the assumption that there are two complementizer positions in the CP: C₁, which bears subject agreement (ϕ) features (and

⁶The version of the paper that I am commenting on here is the one accessible online at <http://www.people.fas.harvard.edu/~szczeg/papers/thattraceweb.pdf>. The version published in MITWPL does not have the section with the account of the Russian facts.

possibly a wh-feature); and C_2 , located higher in the structure, which does not bear these features. Languages differ in whether they have a complementizer of the C_1 type, C_2 type, or both. Thus, English has both: *that* is of type C_2 , while a null complementizer is of type C_1 .

Furthermore, Szczegielniak (1999) argues that intermediate steps of successive-cyclic wh-movement need not be feature-driven. Movement proceeds automatically via the edges of phases on the way – a mechanism Szczegielniak terms “phase hopping”; the only condition required is that at the start of movement, the wh-element should independently move to a phase edge. Objects raise to Spec,vP to check their agreement and case features, and are thus “ready” for long-distance movement. However, subjects are located at Spec,TP which is not a phase edge. So subject wh-words can only be fronted if they are first moved to Spec,CP. An agreeing complementizer is needed to attract them there. This is why subjects in English can be extracted across the null complementizer (agreeing) but not across *that* (non-agreeing).

As for Russian, which does not allow extraction of any kind of element out of a complement clause with an overt complementizer, Szczegielniak argues that it also has two complementizer positions. However, he claims the higher C_2 P phrase, which corresponds to the overt complementizer *čto*, is not a phase at all, and so it blocks movement of any element.

As for the empty complementizer in Russian, Szczegielniak (1999) mentions the possibility that it is an agreeing complementizer of type C_2 . Still, he considers the issue unclear, because, according to him, ungrammatical examples of extraction across a null complementizer can be constructed, such as (35). According to Szczegielniak, such examples are necessary to control for the parenthetical reading of the complementizer-less constructions.

- (35) * *čto* John sporil s nami (*čto*) krugloe?
 What John argued with us (that) round
 ‘John argued with us (that) what is round?’ (Szczegielniak, 1999, p. 19, (33))

5.2.3 Criticism of the existing accounts

Comrie (1971, 1973) brings to light an interesting set of data; however, his solution to the problem follows the methodology of Extended Standard Theory and is acceptable within its framework, but not in light of the contemporary views. The same can be said about the Subjacency-based accounts in (Greenberg, 1988, Meyer, 1997). In addition, the conclusions of (Meyer, 1997) are questionable because of the delicate judgements and marginal examples that he appeals to. First, as he himself admits, sentences with the split conjunction *čtoby*, like (27), are marginal. It seems that the contrast between (28a) and (28b) is due to the marginality of the

split complementizer in (28a) rather than to their different status as regards extraction. Besides, Meyer's model draws heavily on the impossibility of extraction across the complementizer *čto* in cases when it is not embedded under a volitional verb, as in (24). Importantly, however, examples like (23) and (24) always involve either overt negation or a verb with negative semantics, like *doubt*, which means that they are negation islands (cf. (7b)-(7c)); hence, extraction out of these must be blocked for independent reasons.

(Bailyn, 1992) provides an account for the Russian locality phenomena in a footnote only, and does not elaborate on it, so no serious evaluation can be given. For example, Bailyn himself notes that it is not clear why the English *that* and the Russian *čtoby* should be deleted at LF, but not the Russian *čto*. Both *that* and *čto* can be dropped in overt syntax, in contrast to *čtoby*, which can never be omitted, so it is hard to see what unites *čtoby* and *that*.

The accounts of (Stepanov & Georgopoulos, 1997) and (Stepanov, 2001) are also problematic. The problem with (Stepanov & Georgopoulos, 1997) is the idea of the +/-R marking of clauses, which seems ad hoc. The analysis is abandoned in Stepanov's subsequent work in favour of the NP-shell hypothesis; however, the latter also runs into several empirical problems. First, the pronominal correlative *to* is not a frozen form; it is a pronoun that can be inflected for case. The form *to* is accusative (or nominative in other contexts), but the pronoun can also appear in other forms – in fact, it mostly occurs not in accusative but in genitive and other oblique cases, as well as with prepositions (see section 1.4.2). For instance, when the matrix predicate is negated, the correlative cannot stay the same; it has to change its case to genitive (a case of the Russian “genitive of negation”). Stepanov cites the verb *utverždat'* ‘to claim’ as one of the predicates that take an overt *to* correlative:

- (36) ?? Pëtr utverždal to, čto Ivan ljubit Mašu.
 Petr claimed it that Ivan loves Masha
 ‘Petr claimed that Ivan loves Masha’ (grammaticality judgement mine – O.K.)

This sentence is in fact marginal; it can only be acceptable with a contrastive stress on the correlative. However, if we negate the sentence, we get the definitely acceptable (37):

- (37) Pëtr ne utverždal (togo), čto Ivan ljubit Mašu.
 Petr not claimed (it-GEN) that Ivan loves Masha
 ‘Petr did not claim that Ivan loves Masha’

Importantly, subjunctive *čtoby*-clauses can also take the pronominal correlative. The difference is that volition verbs, which introduce subjunctive clauses, do not take the correlative in accusative case – they need genitive, since they assign genitive case:

- (38) a. *Ja xoču to, čtoby Petja počinil mašinu.
 I want it-ACC that Petya would-fix car
 ‘I want Petya to fix the car’
- b. Ja xoču togo, čtoby Petja počinil mašinu.
 I want it-GEN that Petya would-fix car
 ‘I want Petya to fix the car’

These facts undermine Stepanov’s logic in drawing the distinction between indicative and subjunctive complements: if subjunctive complements can also be introduced by a correlative, it follows that they should also be NP-shells, and consequently should not differ from indicative complements as far as extraction is concerned.

Furthermore, Stepanov’s account patterns with Bailyn’s (1992) account in suggesting that Russian complement clauses are structural adjuncts – either by nature (Stepanov, 2001), or after undergoing movement due to the case resistance principle (Bailyn, 1992). If we follow this logic, extraction out of complement clauses must be as ungrammatical as extraction out of “regular” adjuncts, since the two types of sentences are structurally analogous. In fact, this is not true – as already noted in section 5.2.1, extraction from an adjunct is much more severely degraded (compare (39a) with (39b), repeated from (14) on page 136); extraction out of a complex NP also sounds much worse, as shown by (39c).

- (39) a. ?* Kogo Ivan dumaet, čto Marija ljubit?
 Who Ivan knows that Marija loves
 ‘Who does Ivan know that Marija loves?’
- b. * Kogo Maša zaplakala, kogda Petja udaril?
 Whom Masha cried after Petya hit
 ‘*Who did Masha cry after Peter hit?’
- c. * Kogo Ivan videl devušku, kotoraja ljubit _?
 Whom Ivan saw girl which loves
 ‘Who did Ivan see a girl who loves?’

While (39a) is marginal or unacceptable for many speakers, (39b) and (39c) are completely out. Stepanov’s theory does not account for this contrast.

Stepanov’s account would also predict that the presence of an overt correlative in the structure does not influence the grammaticality of the sentence, since structurally the correlative is always there. In fact, extraction across an overt correlative is significantly more degraded:

- (40) a. ?* Kogo Pëtr utverždal, čto Ivan ljubit?
 Who Petr claimed that Ivan loves
 ‘Who did Petr claim that Ivan loves?’

- b. ** Kogo Pëtr utverždal to, čto Ivan ljubit?
 Who Petr claimed it that Ivan loves
 ‘Who did Petr claim that Ivan loves?’

Extraction out of a subjunctive clause across a pronominal correlative is also severely degraded:

- (41) a. Čto ty xočeš, čtoby Petja počinil?
 What you want that Petya would-fix
 ‘What do you want Petya to fix?’
 b. ** Čto ty xočeš togo, čtoby Petja počinil?
 What you want it-GEN that Petya would-fix
 ‘What do you want Petya to fix?’

To sum up, the differences in grammaticality outlined above present problems for both Bailyn’s and Stepanov’s accounts.

Finally, Szczegielniak (1999) argues that long-distance extraction in Russian is prohibited because CP is not a phase. This claim, however, does not follow from Szczegielniak’s own theory. If CP is not a phase, that means subjects cannot be extracted to the phase edge and undergo subsequent cyclic movement. Objects, on the other hand, are not hampered by this problem: they still sit at the vP edge and thus can undergo phase hopping to the next phase edge and upwards; if CP is not a phase, that simply means that they will hop on to the next higher vP edge. So Szczegielniak’s claim predicts that only subjects must be blocked by the Russian *čto*, although Szczegielniak aims to block object movement as well.

As for the example (35), which is introduced to control for a parenthetical reading, it is indeed ungrammatical, with a complementizer or without. However, it is in fact ruled out for independent reasons: the corresponding sentence without extraction is at best marginal.

- (42) ?? Ivan sporil s nami, čto zemlja kruglaja.
 Ivan argued with us that Earth round
 ‘Ivan argued with us that the Earth is round’

In addition, the English counterpart of (35) is out as well, which is probably due to the fact that *argue* is not a “bridge” verb.

- (43) * What did John argue with us (that) is round?

A note on the existence of parenthetical readings is due here. The claim that the Russian sentences with the complementizer *čto* omitted are parenthetical constructions is a widespread one, but there is evidence that it is not correct. First, it is useful to look at the research on parentheticals in German, carried out e.g. in (Reis, 1995). Reis argues that embedded V2 clauses in German are in fact not embedded clauses at all, but matrix clauses with a parenthetical inserted into them. Thus, they should be analysed as (44b), and not as (44a):

- (44) a. [Wo glaubst du, [t' wohnt sie t seit 1985]]?
 Where think you lives she since 1985
 'Where do you think she lives since 1985?' (Reis, 1995, p. 45, (2a))
- b. [Wo [glaubst du], wohnt sie seit 1985]? (Reis, 1995, p. 45, (3a))

A central argument in Reis's paper is the fact that parentheticals of this kind can occur postverbally, as illustrated in (45) (Reis's (4)-(6) on p. 46); Reis argues that in this case an extraction analysis is hardly applicable, so the constructions in question are bona fide parentheticals.

- (45) a. Wo (/In Bonn) wohnt sie meint er (/sagt Paul) mit dem
 Where (/in Bonn) lives she thinks he (/says P.) with the
 Kind seit 1985?
 kid since 1985
 'Where does he think that she has been living with her kid since 1985?'
- b. Wo (/In Bonn) wohnt sie mit dem Kind meint er (/sagt
 Where (/in Bonn) lives she with the kid thinks he (/says
 Paul) seit 1985?
 P.) since 1985
- c. Wo (/In Bonn) wohnt sie mit dem Kind seit 1985 meint
 Where (/in Bonn) lives she with the kid since 1985 thinks
 er (/sagt Paul)?
 he (/says P.)

However, in Russian sentences with a postverbal parenthetical are degraded:

- (46) a. ?? Gde ona živet so svoim rebenkom, ty sčitaeš?
 Where she lives with own child you believe
 'Where do you believe she lives with her child?'
- b. * Gde ona živet, ty sčitaeš, so svoim rebenkom?
 Where she lives you believe with own child

Thus, this argument of Reis's does not work for Russian, showing that the Russian construction is at least distinct from the German one.

An even more important consideration comes from the SOT facts. As noted in section 3.2.2, present tense in Russian can receive a "shifted" interpretation only when it is in the scope of an attitude operator. So, in (47), repeated from (42c) on page 72, this reading is not obtained because the attitude operator is embedded in an adjunct PP.

- (47) Po (prošlogodnemu) utverždeniju Maši, ona ljubit Ivana.
 By (last-year) claim Masha-GEN she loves Ivan.
 'According to Masha's claim (made last year), she loves Ivan.'

But in sentences like (48a)-(48b), where the complementizer *čto* is omitted, the reading is available, just as it is in their counterparts with an overt complementizer.

- (48) a. Ivan dumal, Maša boleet.
Ivan thought Masha is-ill
'Ivan thought that Masha was ill'.
b. Kogo, ty dumal, Maša ljubit?
Whom you thought Masha loves
'Who did you think that Masha loved?'

I conclude that the sentences without the complementizer *čto* display genuine embedding and are not parenthetical constructions.

To sum up, while making many important observations, the theories available in the literature so far do not provide a satisfactory explanation for the locality facts in Russian. In the next section, I will propose my own account of the data.

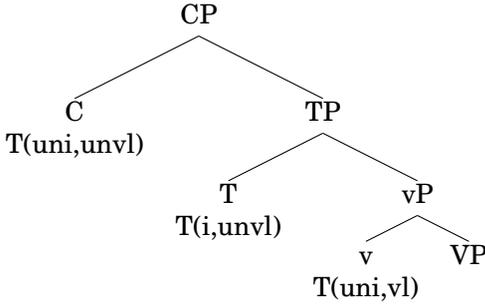
5.3 Long-distance movement in Russian: The account

5.3.1 Phases in Russian and successive-cyclic movement

Let me return briefly to the discussion of Phase Theory and the properties of the English C-T system. As discussed in section 2.1.2, (Chomsky, 2005) provides arguments for the following phase architecture of English: CP and v^*P are phases, while TP and VP, respectively, are dependent on them and get their features derivatively from them. In particular, the temporal and ϕ -features of T are only present there because they are “inherited” from C. Moreover, the fact that T can attract a subject to its Spec is also derivative. Normally, according to Chomsky (2005), only phase heads can be targets of movement (Internal Merge). Thus, in Chomsky’s theory CP and TP form a whole: CP is the phase, Spec,CP is the leftmost edge of the phase, C is the phase head, and T together with its Spec is a domain to which some of the “phasal” properties are transmitted by inheritance.

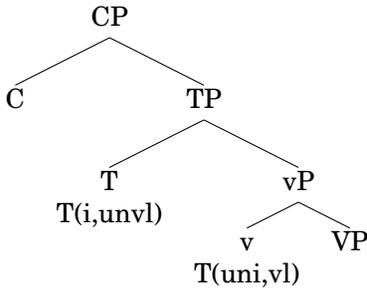
An implementation of the idea of an intrinsic connection between C and T is provided by (Pesetsky & Torrego, 2001, et seq.) (see section 2.1.3). PT propose that C always bears an uninterpretable, unvalued instance of the T(ense) feature. This feature then gets its value and interpretation from the T head (which, in turn, gets its value from v). Schematically, this can be represented as (49), repeated here from (21) on page 36.

(49)



In the previous chapters, I proposed that this analysis does not apply to Russian: namely, in Russian, there is no instance of a T feature on C. The structure for Russian will then look as follows:

(50)



Now an important question has to be addressed: what does this all mean for the phase structure of Russian? If the T feature on C in English is a manifestation of the close connection between C and T, then the evidence for such a connection is absent in Russian.

I propose that the phase architecture of Russian is indeed distinct from that of English, although it is not entirely different. What the two languages have in common is the CP-TP “phasal complex”; in both languages, CP is a phase, and Spec,CP is the left edge of this phase. However, I propose that in Russian, the phase head is T and not C. This means that T does not get any features derivatively; rather, its temporal and ϕ -features are present on it to begin with. In English, the T feature is passed down from C to T, with C retaining an uninterpretable instance of the T feature. But in Russian, C does not inherit any features from T, simply because it is not in the c-command domain of T. It follows that no temporal feature can possibly be present on C.

If the Russian T head does not depend on C, then in Russian C and T should display independent behaviour, not attested in English. In fact, such behaviour can be demonstrated. In English, there is a strict one-to-one correspondence between the form of the complementizer and the finiteness of the verb. The complementizer *that* can only head a finite clause, while a non-finite one either lacks a CP or is headed by *for*:

- (51) a. I think [_{CP} that [_{TP} John came]].
 b. I want [_{TP} John to come]].
 c. [_{CP}For [_{TP} John to come]] would be very good.

In Russian the situation is different. There is no special complementizer for non-finite clauses. The complementizer *čto* heads finite clauses, and it also heads infinitival clauses in combination with the particle *by*, as shown in the examples from (24) on page 10, repeated below as (52b)-(52d).⁷

- (52) a. Ja dumaju, čto Ivan ljubit Mašu.
 I think that Ivan loves Masha
 b. Ja xoču kupit' xleba.
 I want to-buy bread.
 c. Ja pošla v magazin (čtoby) kupit' xleba.
 I went to shop (that-BY) to-buy bread
 'I went to the shop (in order) to buy bread'
 d. *(Čtoby) kupit' xleba, nužno pojti v magazin.
 that to-buy bread necessary to-go to shop
 'In order to buy bread, one needs to go to a shop'

These examples indicate that C by itself does not play a role in establishing the finiteness of the clause.

Now I turn to the peculiar property of the Russian C discussed in this chapter: Russian indicative complement clauses with the complementizer *čto* are opaque for extraction. I propose that this is also due to the way the Russian phase structure is organized. Recall that a crucial point in Phase Theory is that only phase heads can be targets for movement. Thus, in English, only C can be an attractor, while T has a limited ability to attract which it inherits from C. In Russian, on the other hand, the phase head is T. It follows that T is the only element capable of being a target for movement. C, which is not in the domain of T, does not inherit any attracting properties.

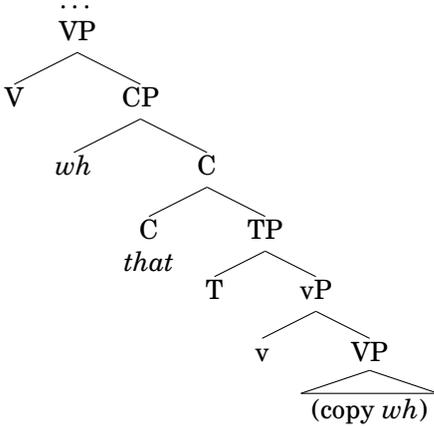
Another vital property of phases is that only the edge of a phase is accessible to subsequent operations after the phase has been completed. This means that in case of long-distance movement, a wh-phrase (or another element moving upwards) has to move successive-cyclically through phase edges. Following Rizzi (2004) I propose that "criterial" (or operator) features are assigned to the intermediate phase heads to make this movement possible. Thus, in English, a wh-phrase extracted out of a complement clause has to move to the edge of the embedded CP, where it will be visible for further movement steps.

The phase structure of Russian creates an obstacle for this process. If the phase head is T, it is also the only element that can attract, so the

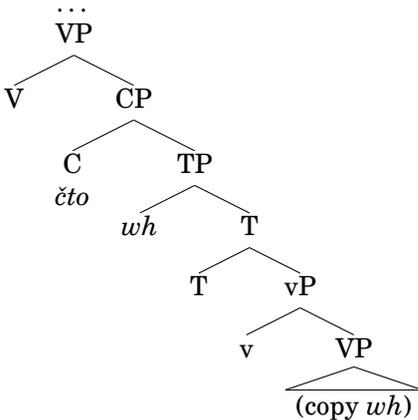
⁷For a detailed analysis of infinitival clauses headed by *čtoby*, see (Junghanns, 1994).

intermediate operator (*wh*) feature is also added to it. However, an element attracted to the edge of TP is not located at the phase edge, since the phase edge is the Spec of CP. So a *wh*-phrase which is extracted out of a complement clause in Russian cannot be visible for subsequent movement after it reaches the intermediate phase. The difference between English and Russian is demonstrated in (53) and (54):

(53) English:



(54) Russian:



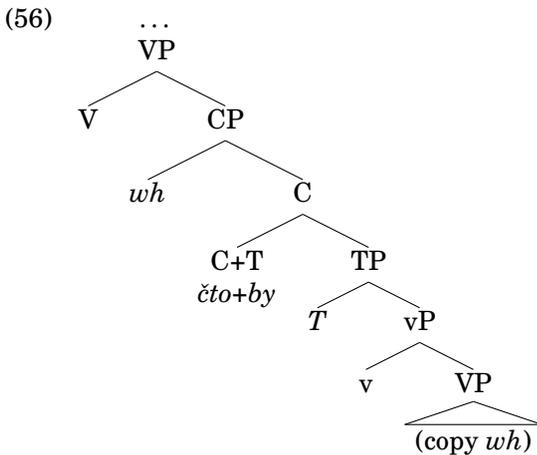
This is why long-distance movement in Russian cannot reach out of a complement clause. However, not all complement clauses in Russian behave alike: recall that extraction out of complement clauses without a complementizer, as well as extraction out of subjunctive clauses with the complementizer *čtoby*, is licit. First, I turn to *čto*-less clauses. What their behaviour indicates is that the notion of phase edge has to be understood in phonological terms. Only if C contains overt material does its Spec count

as the phase edge; if it is empty, then the Spec of TP can function as if it were the leftmost edge of the phase.⁸

Note that even though Spec,TP behaves as the edge of the phase in *čto*-less clauses, it is possible to show that CP still must be projected in these cases. If the complementizer *čto* is omitted, the temporal interpretation of the embedded clause does not change; that is, past embedded under past is normally interpreted as past-shifted, and a simultaneous interpretation is available only in a limited number of contexts. So the meaning of (55) is the same whether or not the complementizer is present. This indicates that C, which “blocks” the temporal agreeing chain (as discussed in Chapter 4), is always present when complements are formed.

- (55) Ivan dumal, (*čto*) Maša bolela.
Ivan thought (that) Masha was-ill

As for the subjunctive complements, the situation with them is easy to explain. I propose that in this case, T is merged together with C, as demonstrated by the form of the complementizer (*čto + by*), so the *wh*-phrase attracted by T actually ends up at the edge of the phase and is then allowed to move on.



This account has a number of advantages. First, it shows that the explanation of the Russian extraction data lies in a general grammatical property of Russian, and introduces a simple distinction between Russian and English. Second, the explanation is different from the analysis of island constraints, such as subject and adjunct islands, complex NP islands, etc. Theories dealing with islands do not have to be concerned with Russian complement clauses now.

⁸Cf. the approach to parsing in (Mulders, 2002), who argues that “the parser does not postulate phonologically empty functional elements, unless overt evidence has been encountered that shows that these functional elements must be present” (p. 6).

In section 5.2.1 I noted that the examples involving extraction out of *čto*-clauses are not as severely degraded as island violations, and that there is some speaker variation. These facts can now also be accounted for. Obviously, no matter how island constraints are derived, the prohibition against extracting anything out of a CP is “added” to all constraints that involve extraction out of embedded clauses. In addition, adjunct island constraint appears to be graver than the prohibition against the extraction out of CP. According to Chomsky (2005), extraction out of adjuncts is impossible because adjuncts are merged late, so that the moved element simply cannot be retrieved from the adjunct clause at the time when it should be merged to the Spec of the higher CP, because the adjunct clause is not there yet. This contrasts with extraction out of complements, where no timing issue is involved: the element to be extracted is there but cannot be reached by the probe.

Some speakers, however, report complete ungrammaticality of extraction across *čto*. This variation can be accounted for if we propose that the operator (*wh*) feature assigned to the intermediate phase head T is for some speakers simultaneously assigned to C, since the whole CP is the phase. In that case, the instance of the operator feature on C will remain unchecked, because the *wh*-phrase can only move as far as Spec,TP and C cannot attract it, not being the phase head. As a result, the derivation crashes.

Another question that arises is where a *wh*-word moves to in a matrix clause. As far as I can tell, there is no evidence against the conclusion that it does not move further than the matrix TP. In fact, examples can be found in which the *wh*-word is lower than the subject, especially when the subject is expressed by a pronoun, as shown in (57):

- (57) *Vy kuda idete?*
 You where go
 ‘Where are you going?’

5.3.2 Remaining issues

The “*wh*-scope marking” construction in Russian

In section 5.2.1, I mentioned that the fully grammatical counterpart of the English long-distance *wh*-questions in Russian is a type of *wh*-scope marking construction, with the *wh*-word *kak* “how” in the matrix clause. This was illustrated by (15) on page 136, repeated here as (58):

- (58) *Kak ty dumaeš, kogo Ivan priglasil?*
 How you think whom Ivan invited
 ‘Who do you think Ivan invited?’

For the sake of completeness, I will say a few words about this construction as well. In the literature, it has been discussed as an instance of a wh-scope marking construction. Wh-scope marking can be defined as a long-distance wh-question where the question word moves no further than the edge of the embedded clause, while the matrix clause contains another wh-word – a so-called wh-scope marker, usually a word meaning “what”. This construction is found in languages as diverse as German, Hungarian, and Hindi, as well as several others. Its properties are slightly different in each of these languages, however there are enough similarities to consider wh-scope marking a uniform phenomenon. Examples from German and Hungarian are given in (59) (Horvath, 1997, p. 510, (1)-(2)).

(59) a. German:

Was glaubt Hans, mit wem Jakob jetzt spricht?
 what believes Hans with whom Jakob now speaks

‘With whom does Jakob think that Hans is now talking?’

b. Hungarian:

Mit gondolsz, hogy kit látott János?
 what-ACC think-2sg that who-ACC saw-3sg John-NOM

‘Who do you think that John saw?’

Several accounts have been offered to explain the structure of wh-scope marking (see (Lutz, Müller, & Stechow, 2000) for an extensive discussion). The two main approaches to the problem have been termed the Direct Dependency Approach and the Indirect Dependency Approach. The Direct Dependency Approach (McDaniel, 1989, a.o.) argues that the wh-scope marker is an expletive base-generated in the matrix Spec,CP which forms a chain with the wh-word in the embedded clause resembling an expletive-associate A-chain. The Indirect Dependency Approach (Dayal, 1994, a.o.) maintains that there is no syntactic chain-like dependency between the two wh-words; rather, the wh-scope marker is a logical operator which quantifies over propositions, and the embedded question functions as a restriction for this operator.

Stepanov (2001) argues that Russian sentences like (58) should be analysed as an instance of wh-scope marking, and offers an analysis in terms of the Indirect Dependency Approach. However, other authors (Gelderen, 2001; Fanselow, 2003) doubt that the Russian *kak*-construction can qualify as genuine wh-scope marking. The alternative is to consider these sentences as a combination of two matrix clauses (or as an idiomatic construction). The most compelling arguments for that are the following:

First, Russian can only use a very limited number of verbs in this construction. In fact, *dumat’* “think”, *sčitat’* “reckon”, *polagat’* “consider”, *kazat’sja* “seem” appear to form an exhaustive list. Although this is also the case for German, where only certain verbs can appear in the wh-scope

marking construction, in German the lexical choices are not so limited; for instance, modal verbs can be used in the construction (as in (60a)). In Russian, this is not allowed (60b).

- (60) a. Was soll Hans glauben, wen Jakob gesehen hat?
 What should Hans think whom Jakob seen has
 'Who should Hans think that Jakob saw?' (McDaniel, 1989, p. 585, (40))
- b. *Kak Ivan mozhët podumat', kogo Jakov videl?
 How Ivan can think whom Yakov saw
 'Who can Ivan think that Yakov saw?'

In addition, the Russian *kak*-construction seems to be used mostly for second person questions; when the third person is used in the first part of the question, it sounds awkward:

- (61) ?Kak думаet Ivan, kogo ljubit Maša?
 How thinks Ivan whom loves Masha
 'Who does Ivan think that Masha loves?'

Second, in other languages, for instance in German, the whole construction can be embedded. This is impossible in Russian; compare the German (62a) with the Russian (62b).

- (62) a. Ich weiss nicht was Hans glaubt mit wem Jakob jetzt
 I know not what Hans thinks with whom Jakob now
 spricht.
 speaks
 'I don't know who Hans thinks that Jakob is talking to now.'
 (McDaniel, 1989, p. 570, (11a))
- b. *Ne znaju, kak думаet Vanja, s kem Petja govorit.
 Not know how thinks Vanya with whom Petya talks
 'I don't know who Vanya thinks that Petya is talking to'

Third, in other languages, the construction can be iterated, with the *wh*-scope marker repeated in several consecutive clauses. Again, in Russian this is impossible:

- (63) a. Was glaubst du, was Hans meint, mit wem Jakob
 What think you what Hans believes with whom Jakob
 gesprochen hat?
 spoken has
 'Who do you think that Hans believes that Jakob spoke to?'
 (McDaniel, 1989, p. 575, (25c))

- b. ?? *Kak ty думаеш, как считает Ваня, с кем Яков
How you think how believes Vanya with whom Яков
поговорил?
talked*

‘Who do you think that Vanya believes that Yakov talked to?’

An important test used to show whether or not the two parts of the construction constitute one sentence is the variable binding test: if an operator in the first part can bind a variable in the second part, this clearly indicates that the second part is embedded in the first. The German “was-w” construction passes the test (Sternefeld, 2002):

- (64) *Was hat jeder_i gesagt, wen er_i heiraten will?
What has everyone said who he marry wants
‘Who did everyone say that he wants to marry?’ (Sternefeld, 2002,
(24b))*

Stepanov (2001) claims that this test can be successfully applied to Russian. He quotes the following example:

- (65) *Kak считает [каждый из студентов]_i, куда ego_i могут отправить?
How thinks each of students where him they-can send
‘Where does each of the students think that they can send him?’
(Stepanov, 2001, p. 165, (9))*

Indeed, the coindexation reading seems possible in this context, though not absolutely natural. However, Stepanov also claims that the same reading is impossible if we replace *kak* by *čto* “what”, a construction which is usually not viewed as an instance of wh-scope marking:

- (66) * *Čto сказал [каждый из студентов]_i? Куда ego_i отправят?
What said each of students where him they-will-send
‘Where did each of the students say that they will send him?’
(Stepanov, 2001, p. 170, (32), grammaticality judgement his)*

In fact, the required reading is not so bad as Stepanov claims, and becomes better if we replace the full stop between the clauses with a comma, making it one sentence intonationally. In that case, the intended interpretation seems to be available as easily as in (65).

However, this does not necessarily mean that we get variable binding in these cases. The problem seems to be with the Russian quantifier *каждый* “every, each”, whose properties are different from those of the English *every*. This crosslinguistic contrast is addressed in (Avrutin & Wexler, 1992). They give the following two interpretations for the Russian example (67) (their (17), p. 270):

- (67) *Každyj student pišet sočinenija.*
 Every student writes papers
- a. Every student is such that he/she writes papers.
 - b. Student(s) write/are writing papers, and it is true of each of them.

For the second interpretation, the presupposition must be that there are some students writing papers, and the sentence is true if every single one of them does so. The first reading is generic, meaning approximately that all students in the world have to write papers and it is simply a property of being a student. Avrutin and Wexler (1992) claim that for the second reading, a relevant set of students must be previously specified in the discourse in an overt way. Otherwise, if the generic reading is unavailable, the sentence becomes odd. In this respect *každyj* is different from the English *every*, which of course can also give rise to these two interpretations, but no particular set of objects has to be specified for the second reading to be available.

In other words, the Russian *každyj* has to be D-linked. For this reason, a “quasi”-bound variable reading can sometimes be available for it, which means that in fact the antecedent of the pronoun is not the quantified NP itself but members of the set introduced before. According to my intuitions as a speaker, even if the set is not introduced overtly in previous discourse (if the sentence is isolated), it can be reconstructed by the hearer since the sentence becomes odd without such a set being specified. In this way, “quasi”-bound variable readings can be forced by Russian speakers relatively easily: (68a) is ungrammatical, as expected, but (68b), where there is a stronger semantic link between the two clauses, is much better.

- (68) a. ?* *Každyj student_i dolžen sdat' èkzameny. K tomu že, emu_i*
 Every student must take exams. Besides to-him
nužno zaščitit' kursovuju.
 necessary to-defend course paper
 lit. 'Every student must take exams. He also has to defend a course paper'
- b. ? *Snačala každyj student_i dolžen sdat' èkzameny, a potom*
 First every student must take exams and then
emu_i eščë pridětsja zaščitit' kursovuju.
 to-him also will-be-necessary to-defend course paper
 lit. 'First every student must take exams, and then he will also have to defend a course paper'

To be sure, instances of such readings are also found in English, such as the famous example from (Sells, 1985):

- (69) Each degree candidate_i walked to the stage. He_i took his_i diploma from the dean and returned to his_i seat.

However, in Russian such examples seem particularly easy to construct. All this casts doubt on the validity of the bound variable test with the word *každyj* for the Russian *kak*-construction. Another possibility is to use the wh-word *kto* “who” in the first part of the sentence. In that case, the bound variable reading becomes virtually impossible to get. Elaborating on Stepanov’s examples, imagine that there is a group of students waiting for a decision on where each of them is going to be sent (e.g. to do dialectological fieldwork). You could then address this group with (70a), but not with (70b), which can only mean the question is about someone else not included in the group; for the intended meaning you would use (70c).

- (70) a. Kto_i думает, что ego_i могут отправить в Сибирь?
Who thinks that him they-can send in Siberia
lit. ‘Who thinks that they can send him to Siberia?’
- b. ?* Kto_i как думает, куда ego_i могут отправить?
Who how thinks where him they-can send
lit. ‘Who thinks that they can send him where?’
- c. Kto как думает, куда вас могут отправить?
Who how thinks where you-pl they-can send
‘What do you think, where can they send you?’

I conclude that the Russian *kak*-construction is better treated as an idiomatic expression than as an instance of wh-scope marking as attested in German and other languages. Exactly how it is analysed should be a matter of further research; it is possible to suggest that certain verbal forms, such as the 2nd person forms of the verbs *думает*, *считает*, etc. can take a question as a complement, although in general it is disallowed by the selectional properties of these verbs; the construction is then saved by the addition of the “scope marker” *kak*.

Extraction in other Slavic languages

A detailed discussion of Slavic languages other than Russian is outside the scope of this thesis, and I will leave this important topic for the future. However, I will mention some relevant data briefly. Judgements on the grammaticality of long-distance extraction data in Polish vary in the literature. Thus, Szczegielniak (1999) claims that extraction from a complement clause is always grammatical in Polish (in contrast to Russian). Some of the Polish speakers that I consulted also allowed extraction. However, most authors who mention the problem (Borsley, 1983; Meyer, 2003; Toman, 1981) argue that extraction out of indicative complements in Polish is ungrammatical. Apparently, inter-speaker variation similar to Russian can be found in Polish. Importantly, (Witkoś, 1995; Giejgo, 1981) claim that extraction out of indicative complements is prohibited in Polish, but extraction out of subjunctive complements is allowed:

- (71) a. *Co Tomek móvi że Maria wie t?
 What Tomek says that Mary knows
 ‘What does Tomek say that Mary knows?’ (Witkoś, 1995, p. 223, (2))
- b. Co Tomek chce żeby mama mu przeczytała?
 What Tomel wants that-cond mum to-him read
 ‘What does Tomek want mum to read to him?’ (Witkoś, 1995, p. 224, (3))

The Polish subjunctive complementizer *żeby* is formed in the same way as the Russian complementizer *čtoby*: it consists of the indicative complementizer *że* combined with the subjunctive particle. Furthermore, Polish is a “non-SOT” language with the same temporal interpretations available for embedded tenses as Russian (see Kusumoto, 1999 for discussion):

- (72) a. Ania powiedziała, że Marcin płacze
 Ania said that Marcin cries
 ‘Ania said that Marcin was crying’
- b. Ania powiedziała, że Marcin płakał
 Ania said that Marcin cried
 ‘Ania said that Marcin had been crying’ (Kusumoto, 1999, p. 86, (123a-b))

Therefore, I tentatively conclude that my analysis of Russian SOT and extraction facts can be extended to Polish.

Czech, on the other hand, is claimed to have unlimited extraction from indicative complement clauses by (Meyer, 2003, Toman, 1981). However, Jakub Dotlacil (p. c.) points out that there is some speaker variation with extraction out of indicative complements in Czech as well, which is reported e.g. in (Štícha, 1996). Extraction out of subjunctive complements is unambiguously grammatical, so the same pattern seems to be observed here; Czech is also a non-SOT languages. Further research will probably be needed to establish whether Czech really patterns with Russian (and Polish).

5.4 Summary

In this chapter I have presented my account of long-distance movement in Russian. My aim has been to show that the absence of a T-feature on C, which leads to the absence of SOT effects, and the ban on extraction out of complements are caused by the same property of Russian grammar: the fact that T and not C is a phase head in this language, in contrast to English where C is the head of the phase. It follows that features of the T head are not shared with C. It also follows that a phrase extracted out of

an embedded clause in Russian cannot be attracted to the phase edge, and further successive-cyclic movement becomes impossible.

Chapter 6

Conclusions

This dissertation deals with two distinct domains of grammar: Sequence of Tense (SOT) and long-distance movement. My starting point was the observation that Russian behaves differently from English in these domains. English is an “SOT language”, displaying specific dependent readings of past tense in complement clauses; Russian is a “non-SOT” language, lacking those readings. Besides, Russian notoriously disallows long-distance extraction out of complement clauses, which is in most cases fully grammatical in English. In this thesis, I show that these two facts stem from one underlying difference between the English and the Russian grammar, and I define this difference in terms of Phase Theory.

In Chapter 1 of this dissertation I introduce the data to be discussed and spend some time on the properties of Russian grammar that are relevant to the subsequent discussion. Chapter 2 is dedicated to the general syntactic background of my research. First, I define the notion of the C-T system and outline the basics of Phase Theory, which argues for a strong connection between C and T. According to Phase Theory as formulated in Chomsky (2005), in English CP is a phase, and TP inherits some phase-like properties from it. In this thesis, I propose that in Russian the relation between the two is different: the phase is also CP, so the leftmost edge of the phase is Spec,CP, but the phase head is T, not C. I go on to present the theory of Pesetsky and Torrego (2001, et seq.) who argue that C bears an uninterpretable T feature, and outline their theory of Agree as feature sharing.

A substantial part of this dissertation deals with Sequence of Tense in English and Russian. There is a long tradition of treating tenses as pronominal expressions; consequently, relations between tenses in a sentence (and in the discourse) are often defined in terms of anaphora, and the apparatus of the Binding Theory has been applied to temporal relations. Since I also adhere to the “pronominal” view of tenses, in Chapter 2 I discuss Binding Theory, the notion of binding and its possible applications to

the theory of tense. Many theories define the relation between the matrix and the embedded tense in purely semantic terms: as a binding relation in the logical syntax sense. I opt for a different strategy and use Reuland's (2001, 2005) theory of binding based on Agree relations to account for SOT interpretations.

Chapter 3 contains an in-depth presentation of problems connected with the SOT theory. First, I present the relevant data and the problems that they pose for the theory of tense. I then discuss possible ways to solve these problems and the existing approaches to the nature of tenses. I give a more detailed analysis of several accounts of SOT, including those that propose semantic theories of SOT and those that place the relation between matrix and embedded tenses into syntax. My own proposal involves a syntactic mechanism of temporal binding, however I argue that the behaviour of tenses is also governed by certain semantic constraints (such as, for example, Abusch's Upper Limit Constraint for tense interpretation).

In Chapter 4, I develop my own account of SOT. I start with SOT languages like English and Dutch, and then demonstrate why Russian is different from them. For SOT languages, I argue that in case of simultaneous interpretation of past embedded under past, the verb in the complement clause enters the derivation with an unvalued T(ense) feature, which gets its value later in the derivation from the higher T, via an Agree chain proceeding through the embedded C which also carries an instance of T feature. As a consequence, a bound interpretation arises. I argue that in Russian such a chain cannot be formed, because C does not bear an instance of a T feature. This follows from the phase structure of Russian: the temporal feature of the Russian T is not inherited from C but present on the T head independently, hence no "traces" of tense are found on C.

In Chapter 5 I tackle the other problem that I set out to solve in this thesis: the lack of long-distance movement in Russian. I start with a brief overview of literature on successive-cyclic movement and island constraints, then present the Russian data, which remain unaccounted for by the major theories of long-distance movement. I also provide a criticism of the existing accounts of extraction in Russian. My own account of the Russian data hinges, again, on the phase structure of this language. I argue that in Russian, elements cannot be easily extracted from embedded clauses because they never get to the edge of the phase: since only the phase head, hence T, can attract, the moved elements do not end up in Spec,CP.

Let me conclude with some prospects for future research. This dissertation was mostly concerned with English (and to some extent Dutch) as opposed to Russian. At the end of Chapter 5, I noted that I expect my analysis to extend to (some) other Slavic languages, at least to Polish and Czech. This, of course, deserves a more careful and detailed study. On the other hand, I did not offer any analysis of Japanese, apart from relying on Kusumoto (1999) in the analysis of Japanese relative clauses. However, at least to mention Japanese is important for me because it is another exam-

ple of a non-SOT language, but with typological properties very different from those of Russian or indeed any Indo-European language. Interestingly, Japanese has long-distance movement, which is not at all limited in the way that it is in Russian. This is not predicted by my analysis, but I can suggest that the Agree chain between T heads that I propose for English can “break” for reasons other than those identified for Russian. Naturally, this problem should be given more attention to in the future.

There are also some aspects of extraction in Russian that were put aside in my discussion. Thus, there is a debate in the literature whether long-distance scrambling obeys the same limitations as long-distance wh-movement (which is virtually the only type of movement that I have considered here). (Müller & Sternefeld, 1993) claim that long-distance scrambling is freer than wh-movement, while (Bailyn, 2003) claims that they are equally limited; the respective judgements are presented in (1a)-(1b) and (1c)-(1d).

- (1) a. Mne Katju_i kažetsja, [čto otpustit' t_i odnu tak
to-me Katya-ACC seems that to-let-go alone so
pozдно] bylo by bezumiem.
late be SUBJ insanity
'It seems to me that it would be insane to allow Katya to go
alone so late at night' (Müller & Sternefeld, 1993, p. 467,
(9a))
- b. ?* Kogo_i tebe kažetsja, [čto otpustit' t_i odnu tak pozдно]
whom to-you seems that to-let-go alone so late
bylo by bezumiem?
be SUBJ insanity
'Who do you think that it would be insane to let go alone so
late at night?' (Müller & Sternefeld, 1993, p. 467, (10a))
- c. * Kogo_i Marina znaet [čto [Ivan ljubit t_i]]?
whom Marina knows that Ivan loves
'Who does Marina know that Ivan loves?'
- d. * Borisa_i Marina znaet [čto [Ivan ljubit t_i]]
Boris-ACC Marina knows that Ivan likes
'Marina knows that Ivan likes Boris.' (Bailyn, 2003, p. 167,
(26a-b))

Note that my account predicts ungrammaticality for all types of long-distance movement in Russian: no phrases can land at the edge of a phase, independent of the type of movement. According to my judgements, long-distance scrambling examples are better than their wh-movement counterparts, but still not completely grammatical. Thus, (1a), taken by Müller & Sternefeld from (Zemskaja, 1973), is rather marginal even for colloquial speech. The improved status of such examples must be due to the fact

that scrambled DPs are more discourse linked than *wh*-phrases. The same explanation can be applied to the fact that in many cases, long-distance relativization is more acceptable than long-distance movement in questions. I leave a more careful examination of these matters for the future.

References

- Abusch, D. (1988). Sequence of tense, intensionality and scope. In *Proceedings of the 7th West Coast Conference on Formal Linguistics* (pp. 1–14). CSLI, Stanford.
- Abusch, D. (1997). Sequence of tense and temporal de re. *Linguistics and Philosophy*, 20, 1-50.
- Abusch, D. (1998). Generalizing tense semantics for future contexts. In S. Rothstein (Ed.), *Events and grammar* (p. 13-33). Dordrecht: Kluwer.
- Altshuler, D. (2004). A simultaneous perception of things: SOT in Russian. *Snippets*, 8, 5-6.
- Altshuler, D. (2005). *Sequence of tense and double access*. (Ms., Rutgers University)
- Aoun, J., & Sportiche, D. (1981). On the formal theory of government. *The Linguistic Review*, 2, 211–236.
- Avrutin, S., & Wexler, K. (1992). Development of principle B in Russian: Coindexation at LF and coreference. *Language Acquisition*, 2, 259–306.
- Babyonyshev, M., & Matushansky, O. (2006). Back to the past: Russian past tense revisited. In J. Lavine, S. Franks, M. Tasseva-Kurktchieva, & H. Filip (Eds.), *Proceedings of FASL 14: The Princeton Meeting*. Ann Arbor, Michigan: Michigan Slavic Publications.
- Bailyn, J. F. (1992). LF movement of anaphors and the acquisition of embedded clauses in Russian. *Language Acquisition*, 2(4), 307-335.
- Bailyn, J. F. (1995a). *A configurational approach to Russian 'free' word order*. Unpublished doctoral dissertation, Cornell University.
- Bailyn, J. F. (1995b). Underlying phrase structure and 'short' verb movement in Russian. *Journal of Slavic Linguistics*, 3(1), 13-58.
- Bailyn, J. F. (2003). Does Russian scrambling exist? In S. Karimi (Ed.), *Word order and scrambling* (pp. 156–176). Malden, MA: Blackwell Publishers.
- Bailyn, J. F. (2006). Against the scrambling anti-movement movement. In J. Lavine, S. Franks, H. Filip, & M. Tasseva-Kurktchieva (Eds.), *Formal Approaches to Slavic Linguistics 14: The Princeton meeting*. Ann Arbor, MI: University of Michigan Slavic Publications.

- Barentsen, A. (1996). Shifting points of orientation in Modern Russian: Tense selection in 'reported perception'. In T. A. Janssen & W. van der Wurff (Eds.), *Reported speech* (pp. 15–55). Amsterdam: John Benjamins.
- Boeck, W. (1957). Zum Tempusgebrauch des Russischen in Objekt- und Subjektsätzen. *Zeitschrift für Slavistik*, 2, 206–218.
- Boeck, W. (1958). Der Tempusgebrauch in den russischen Objekt- und Subjektsätzen, seine historische Entwicklung und sein stilistischer Wert. *Zeitschrift für Slavistik*, 3, 209–234.
- Bondarko, A. V. (1983). *Principy funkcional'noj grammatiki i voprosy aspektologii [Principles of functional grammar and issues of aspectology]*. Leningrad: Nauka.
- Bondarko, A. V. (Ed.). (1987). *Teorija funkcional'noj grammatiki. Vvedenie. Aspektual'nost'. Vremennaja lokalizovannost'. Taksis [Theory of functional grammar. Introduction. Aspectuality. Temporal localisation. Taxis]*. Leningrad: Nauka.
- Bondarko, A. V. (1990). O značenii vidov russkogo glagola [About the meanings of the Russian verbal aspects]. *Voprosy jazykoznanija*, 4, 5–24.
- Borik, O. (2002). *Aspect and reference time*. Doctoral dissertation, Utrecht University.
- Borsley, R. (1983). A note on preposition stranding. *Linguistic Inquiry*, 14, 338–342.
- Bošković, Ž. (2001). On the interpretation of multiple questions. In *Linguistic variation yearbook*. Amsterdam: John Benjamins.
- Bošković, Ž. (2002). On multiple wh-fronting. *Linguistic Inquiry*, 33, 351–384.
- Brecht, R. (1977). Čtoby or čto and by. *Folia Slavica*, 1, 33–41.
- Brody, M. (1997). Perfect chains. In L. Haegeman (Ed.), *Elements of grammar* (p. 139–167). Dordrecht: Kluwer Academic Publishers.
- Cable, S. (2005). *Binding local person pronouns without semantically empty features*. (Ms., MIT)
- Chien, Y.-C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, 1(3), 225–295.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232–286). New York: Holt, Rinehart, and Winston.
- Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax*. New York: Academic Press.
- Chomsky, N. (1980). On binding. *Linguistic Inquiry*, 11, 1–46.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.

- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. (1999). Derivation by phase. In *MIT occasional papers in linguistics* (Vol. 18, p. 1-43). Cambridge, MA: MITWPL. (Reprinted in: M. Kenstowicz (Ed.). (2001). *Ken Hale: A Life in Language* (pp. 1-52). Cambridge, MA: MIT Press.)
- Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (p. 89-155). Cambridge, MA: MIT Press.
- Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press.
- Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. 3, p. 104 -131). Oxford: Oxford University Press. (First published in *MIT Working Papers in Linguistics* 20, 2001, p. 1-28.)
- Chomsky, N. (2005). *On phases*. (To appear in: C. P. Otero et al., eds., *Foundational Issues in Linguistic Theory* (MIT))
- Chomsky, N. (2006). *Approaching UG from below*. (Ms.)
- Cinque, G. (1990). *Types of A' dependencies*. Cambridge, MA: MIT Press.
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Comrie, B. S. (1971). *Aspects of sentence-complementation in Russian*. Doctoral dissertation, Cambridge.
- Comrie, B. S. (1973). Clause structure and movement constraints in Russian. In C. Corum, T. C. Smith-Stark, & A. Weiser (Eds.), *You take the high node and I'll take the low node: Papers from the comparative syntax festival* (pp. 291-304). Chicago: Chicago Linguistic Society. (Reprinted in C. V. Chvany and R. D. Brecht (Eds.). (1980). *Morphosyntax in Slavic* (pp. 98-113). Columbus, OH: Slavica)
- Comrie, B. S. (1976). *Aspect*. Cambridge: Cambridge University Press.
- Comrie, B. S. (1985). *Tense*. Cambridge: Cambridge University Press.
- Comrie, B. S. (1986). Tense in indirect speech. *Folia Linguistica*, 20, 265-296.
- Costello, D. P. (1961). Tense in indirect speech in Russian. *The Slavonic and East European Review*, 39, 489-496.
- Dayal, V. S. (1994). Scope marking as indirect wh-dependency. *Natural Language Semantics*, 2, 137-170.
- Doron, E. (1991). Point of view as a factor of content. In S. Moore & A. Wyner (Eds.), *Proceedings from Semantics and Linguistic Theory I* (pp. 51-64). Ithaca, N.Y.: CLC Publications.
- Emonds, J. (1976). *A transformational approach to English syntax*. New York: Academic Press.
- Emonds, J. (1978). The verbal complex V'-V in French. *Linguistic Inquiry*, 9, 151-175.

- Enç, M. (1986). Towards a referential analysis of temporal expressions. *Linguistics and Philosophy*, 9, 405–426.
- Enç, M. (1987). Anchoring conditions for tense. *Linguistic Inquiry*, 18, 633–657.
- Enç, M. (1996). Tense and modality. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 345–358). Oxford: Blackwell.
- Enç, M. (2004). Rethinking past tense. In J. Guéron & J. Lecarme (Eds.), *The syntax of time* (pp. 203–215). Cambridge, MA: MIT Press.
- Erteschik-Shir, N. (1973). *On the nature of island constraints*. Doctoral dissertation, MIT.
- Fanselow, G. (2003). Partial movement. In M. Everaert & H. van Riemsdijk (Eds.), *SYNCOM (The syntax companion)*. The LingComp Foundation (<http://www.uilots.let.uu.nl/syncom>).
- Geis, M. (1970). *Adverbial subordinate clauses in English*. Doctoral dissertation, MIT.
- Gelderen, V. van. (2001). Partial wh-movement in Russian. In T. van der Vouden & H. Broekhuis (Eds.), *Linguistics in the Netherlands* (pp. 89–100). Amsterdam: John Benjamins.
- Giejgo, J. A. (1981). *Movement rules in Polish syntax*. Doctoral dissertation, University College, London.
- Giorgi, A., & Pianesi, F. (1997). *Tense and aspect: From semantics to morphosyntax*. New York and Oxford: Oxford University Press.
- Greenberg, G. R. (1988). Relativization and subadjacency in Russian. In *Proceedings of the Eastern States Conference on Linguistics (5th, Philadelphia, PA, September 30–October 2, 1988)*.
- Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, 24, 69–101.
- Heck, F., & Müller, G. (2000). Successive cyclicity, long-distance superiority, and local optimization. In R. Billerey & B. D. Lillehaugen (Eds.), *Proceedings of WCCFL 19* (pp. 218–231). Somerville, MA: Cascadilla Press.
- Heim, I. (1994). Comments on Abusch's theory of tense. In H. Kamp (Ed.), *Ellipsis, tense and questions. DYANA deliverable R2.2.B* (pp. 143–170). University of Amsterdam.
- Heim, I. (1998). Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. In U. Sauerland & O. Percus (Eds.), *The interpretative tract, MIT working papers in linguistics* (Vol. 25). Cambridge, MA: MITWPL.
- Hinrichs, E. (1986). Temporal anaphora in discourses of English. *Linguistics and Philosophy*, 9, 63–82.
- Horvath, J. (1997). The status of 'wh-expletives' and the partial wh-movement construction of Hungarian. *Natural Language and Linguistic Theory*, 15, 509–572.
- Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Doctoral dissertation, MIT.

- Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2), 231–270.
- Jakobson, R. (1932). Zur Struktur des russischen Verbums. In *Charisteria Gvilelmo Methsio quinquagenario a discipulis et Circuli Linvistici Pragensis sodalibus oblata* (p. 74-83). Prague: Cercle Linguistique de Prague.
- James, F. (1986). *The semantics of the English subjunctive*. Vancouver: University of British Columbia Press.
- Junghanns, U. (1994). *Syntaktische und semantische Eigenschaften russischer finaler Infinitiveinbettungen*. München: Verlag Otto Sagner.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (Vol. 1, pp. 277–322). Amsterdam: Mathematisch Centrum.
- Kamp, H., & Rohrer, C. (1984). *Indirect discourse*. (Ms., University of Texas, Austin, and University of Stuttgart)
- Kaplan, D. (1969). Quantifying in. In D. Davidson & J. Hintikka (Eds.), *Words and objections*. Dordrecht: Reidel.
- Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–614). New York: Oxford University Press.
- Kayne, R. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Khomitsevich, O., & Reuland, E. (2007). On the encoding of temporal dependencies. In P. Costa & L. Schürcks (Eds.), *Linguistic investigations into formal descriptions of Slavic languages (FDSL 6)* (Vol. 1, p. 227-244). Frankfurt am Main: Peter Lang.
- Kondrashova, N. (1999). Barenaked tenses and their morphological outfits. In *Proceedings of NELS 29* (Vol. 1, pp. 183–195).
- Kondrashova, N. (2006). *Is Russian a split-SOT language?* (Paper presented at FASL 14)
- Koornneef, A. W., Wijnen, F., & Reuland, E. (2006). Towards a modular approach to anaphor resolution. In R. Artstein & M. Poesio (Eds.), *ESSLI Ambiguity in Anaphora workshop proceedings* (pp. 65–72).
- Koster, J. (1978). Why subject sentences don't exist. In S. J. Keyser (Ed.), *Recent transformational studies in European languages*. Cambridge, MA: MIT Press.
- Kratzer, A. (1998). More structural analogies between pronouns and tenses. In *The proceedings of Semantics and Linguistic Theory VIII*. CLC Publications.
- Kratzer, A. (2006). *Minimal pronouns*. (Ms., University of Massachusetts at Amherst)
- Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: The MIT Press.
- Kusumoto, K. (1999). *Tense in embedded contexts*. Doctoral dissertation,

- University of Massachusetts.
- Ladusaw, W. (1977). Some problems with tense in PTQ. *Texas Linguistic Forum*, 6, 89-102. (University of Texas, Austin)
- Larson, R. (1988). On the double object construction. *Linguistic Inquiry*, 19, 335-391.
- Lasnik, H., & Saito, M. (1984). On the nature of proper government. *Linguistic Inquiry*, 15, 235-289.
- Lebeaux, D. (1988). *Language acquisition and the form of grammar*. Doctoral dissertation, University of Massachusetts at Amherst.
- Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, 88, 513-543.
- Lutz, U., Müller, G., & Stechow, A. von (Eds.). (2000). *Wh-scope marking*. Amsterdam: John Benjamins.
- Manzini, M. R. (1994). Sentential complementation: The subjunctive. In P. Coopmans, M. Everaert, & J. Grimshaw (Eds.), *Lexical specification and insertion* (pp. 185-209). Hillsdale, NJ: LEA.
- Maslov, J. S. (1984). *Očerki po aspektologii [Outlines of aspectology]*. Leningrad: Izdatel'stvo Leningradskogo Universiteta.
- Maslov, J. S. (1985). *Contrastive studies in verbal aspect in Russian, English, French, and German*. Heidelberg: Groos.
- McCloskey, J. (2002). Resumption, successive cyclicity, and the locality of operations. In S. D. Epstein & D. Seeley (Eds.), *Derivation and explanation in the Minimalist Program* (pp. 184-226). Oxford: Blackwell.
- McDaniel, D. (1989). Partial and multiple wh-movement. *Natural Language and Linguistic Theory*, 7, 565-604.
- Meyer, R. (1997). Extraction from čto- and čtoby- complement clauses in Russian. In U. Junghanns & G. Zybatow (Eds.), *Formale Slavistik* (pp. 165-176). Frankfurt am Main: Vervuert Verlag.
- Meyer, R. (2003). Structural differences in Polish vs. Czech wh-questions. In A. Przepiórkowski & P. Bański (Eds.), *Generative linguistics in Poland 5. Proceedings of the GLiP-5 conference held in Warsaw, Poland, 30 November-1 December 2002* (pp. 63-78).
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221-242). Dordrecht: Reidel.
- Mulders, I. (2002). *Transparent parsing*. Doctoral dissertation, Utrecht University.
- Müller, G. (1995). *A-bar syntax: A study in movement types*. New York: Mouton de Gruyter.
- Müller, G., & Sternefeld, W. (1993). Improper movement and unambiguous binding. *Linguistic Inquiry*, 24, 461-507.
- Murasugi, K. (1991). *Noun phrases in Japanese and English: A study in syntax, learnability, and acquisition*. Doctoral dissertation, University of Connecticut, Storrs.

- Musan, R. (1997). Tense, predicates and life-time effects. *Natural Language Semantics*, 5, 271–301.
- Musan, R. (1999). Temporal interpretation and information-status of noun phrases. *Linguistics and Philosophy*, 22, 621–661.
- Nunes, J., & Uriagereka, J. (2000). Cyclicity and extraction domains. *Syntax*, 3, 20–43.
- Ogihara, T. (1989). *Temporal reference in English and Japanese*. Doctoral dissertation, University of Texas, Austin.
- Ogihara, T. (1995). The semantics of tense in embedded clauses. *Linguistic Inquiry*, 26, 663–679.
- Ogihara, T. (1996). *Tense, attitudes, and scope*. Dordrecht: Kluwer.
- Ogihara, T. (1999a). Double-access sentences generalized. In *The proceedings of Semantics and Linguistic Theory IX* (pp. 224–236). Cornell University, Ithaca, N.Y.: CLC Publications.
- Ogihara, T. (1999b). Tense and aspect. In N. Tsujimura (Ed.), *Handbook of Japanese linguistics* (pp. 326–348). Oxford: Blackwell.
- Ogihara, T. (2003). A scope theory of tense and adnominal modifiers. In S. C. et al. (Ed.), *Empirical and theoretical investigations into language: A festschrift for Masaru Kajita*. Tokyo, Japan: Kaitakusha.
- Padučeva, E. V. (1996). *Semantičeskie issledovanija: Semantika vremeni i vida v rusском jazyke. Semantika narrativa [Semantic investigations: Semantics of tense and aspect in Russian. Semantics of the narrative]*. Moscow: Škola “Jazyki Russkoj Kul’tury”.
- Partee, B. (1973). Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy*, 70, 601–609.
- Partee, B. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, 7, 243–286.
- Perlmutter, D. (1971). *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart, and Winston.
- Pesetsky, D. (1987). Wh in situ: Movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The representation of (in)definiteness*. Cambridge, MA: MIT Press.
- Pesetsky, D., & Torrego, E. (2001). T-to-C movement: Causes and consequences. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (p. 355–426). Cambridge, MA: MIT Press.
- Pesetsky, D., & Torrego, E. (2004a). Tense, case, and the nature of syntactic categories. In J. Guéron & J. Lecarme (Eds.), *The syntax of time*. Cambridge, MA: MIT Press.
- Pesetsky, D., & Torrego, E. (2004b). *The syntax of valuation and the interpretability of features*. (To appear in: S. Karimi, V. Samiian & W. Wilkins (Eds.). *Clever and Right: A Festschrift for Joe Emonds*)
- Pollock, J.-Y. (1989). Verb movement, UG and the structure of IP. *Linguistic Inquiry*, 20, 365–424.
- Prior, A. N. (1967). *Past, present, and future*. Oxford: Clarendon Press.

- Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Philosophical Review*, 53, 177–187.
- Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
- Reinhart, T. (1986). Center and periphery in the grammar of anaphora. In B. Lust (Ed.), *Studies in the acquisition of anaphora* (Vol. 1). Dordrecht: Reidel.
- Reinhart, T. (2000). Strategies of anaphora resolution. In H. Bennis, M. Everaert, & E. Reuland (Eds.), *Interface strategies* (pp. 295–325). Amsterdam: Royal Academy of Arts and Sciences.
- Reinhart, T. (2006). *Interface strategies*. Cambridge, MA: MIT Press.
- Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657–720.
- Reis, M. (1995). Extraction from verb-second clauses in German? In U. Lutz & J. Pafel (Eds.), *On extraction and extraposition in German*. Amsterdam: John Benjamins.
- Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry*, 32(3), 439–492.
- Reuland, E. (2005). Agreeing to bind. In H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz, & J. Koster (Eds.), *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*. Berlin/New York: Mouton de Gruyter.
- Reuland, E. (2007). *Minimal versus full pronouns*. (Ms., Utrecht University)
- Řezáč, M. (2004). *Agree and merge*. (Ms., available at <http://ling.auf.net/lingBuzz/000041>)
- Rizzi, L. (1982). Violations of the wh-island constraint in Italian and the subjacency condition. In L. Rizzi (Ed.), *Issues in Italian syntax* (pp. 49–76). Dordrecht: Foris.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (p. 281–338). Dordrecht: Kluwer Academic Publishers.
- Rizzi, L. (2002). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond. The cartography of syntactic structures* (Vol. 3, pp. 223–252). Oxford: Oxford University Press.
- Rizzi, L. (2004). *On the form of chains: Criterial positions and ECP effects*. (Ms., Università di Siena)
- Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral dissertation, MIT.
- Sarkar, A. (1998). The conflict between future tense and modality: The case of *will* in English. *U. Penn Working Papers in Linguistics*, 5(2), 91–117.
- Schlenker, P. (1999). *Propositional attitudes and indexicality: A cross-categorical approach*. Unpublished doctoral dissertation, MIT.

- Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, 26, 29-120.
- Schlenker, P. (2004). Sequence phenomena and double access generalized: Two remarks on tense, person, and mood. In J. Guéron & J. Lecarme (Eds.), *The syntax of time* (p. 555-595). Cambridge, MA: MIT Press.
- Sells, P. (1985). *Lectures on contemporary syntactic theories*. Stanford: Center for the Study of Language and Information.
- Sharvit, Y. (2003). Embedded tense and universal grammar. *Linguistic Inquiry*, 34, 669-681.
- Slioussar, N. (2007). *Grammar and information structure: A study with reference to Russian*. Doctoral dissertation, Utrecht University.
- Starke, M. (2001). *Move dissolves into Merge: A theory of locality*. Doctoral dissertation, University of Geneva.
- Stechow, A. von. (1995). *Tense in intensional contexts: Two semantic accounts of Abusch's theory of tense*. (Ms.)
- Stechow, A. von. (2003a). *Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers*. (Paper presented at NELS 33)
- Stechow, A. von. (2003b). *Postscript to "Feature deletion under semantic binding"*. (Ms., available at: <http://vivaldi.sfs.nphil.uni-tuebingen.de/arnim10/>)
- Stechow, A. von. (2004). Binding by verbs: Tense, person and mood under attitudes. In H. Lohnstein & S. Trissler (Eds.), *The syntax and semantics of the left periphery*. Berlin: de Gruyter.
- Steele, S. (1975). Past and irrealis: Just what does it all mean? *International Journal of American Linguistics*, 41(3), 200-217.
- Stepanov, A. (1998). On wh-fronting in Russian. In *Proceedings of NELS* (Vol. 28, p. 453-467).
- Stepanov, A. (2001). *Cyclic domains in syntactic theory*. Unpublished doctoral dissertation, University of Connecticut.
- Stepanov, A., & Georgopoulos, C. (1997). Structure building and the conceptual interface: An analysis of Russian long-distance wh-questions. In M. Lindseth & S. Franks (Eds.), *Proceedings of Formal Approaches to Slavic Languages V* (pp. 295-311). Ann Arbor, MI: Michigan Slavic Publications.
- Sternefeld, W. (2002). Wh-expletives and partial wh-movement: Two non-existing concepts? In *Issues in formal German(ic) typology* (pp. 285-305). Amsterdam: John Benjamins.
- Štícha, F. (1996). Křížení vět v češtině. *Naše řeč*, 79, 26-31.
- Stjepanović, S. (1999). *What do second position cliticization, scrambling, and multiple wh-fronting have in common?* Doctoral dissertation, University of Connecticut, Storrs.
- Stowell, T. (1981). *The origins of phrase structure*. Doctoral dissertation, MIT, Cambridge, MA.
- Stowell, T. (1995). What do the present and past tenses mean? In P. M.

- Bertinetto, V. Bianchi, J. Higginbotham, & M. Squartini (Eds.), *Temporal reference, aspect and actionality* (Vol. 1: Semantic and Syntactic Perspectives). Torino: Rosenberg and Sellier.
- Švedova, N. J. (Ed.). (1982). *Russkaja grammatika [Russian grammar]*. Moscow: Nauka.
- Szabolcsi, A., & Dikken, M. den. (1999). Islands. *GLOT International*, 4(6), 3–9. (Reprinted in L. Cheng & R. Sybesma (Eds.). (2003). *The Second GLOT State-of-the-Article Book*. Berlin: Mouton de Gruyter)
- Szczegieliński, A. (1999). “That-trace effects” cross-linguistically and successive-cyclic movement. In K. Arregi, B. Bruening, C. Krause, & V. Lin (Eds.), *Papers on morphology and syntax, cycle one. MIT working papers in linguistics* (Vol. 33, p. 369-393). MITWPiL. (Available at: <http://www.people.fas.harvard.edu/~szczeg/papers/thattraceweb.pdf>)
- Toman, J. (1981). Aspects of multiple wh-movement in Polish and Czech. In R. May & J. Koster (Eds.), *Levels of syntactic representation* (pp. 293–302). Dordrecht: Foris.
- Uriagereka, J. (1999). Multiple Spell-out. In S. Epstein & N. Hornstein (Eds.), *Working minimalism* (pp. 251–282). Cambridge, MA: MIT Press.
- Witkoś, J. (1995). Wh-extraction from clausal complements in Polish: A minimality/locality account. *Folia Linguistica*, 29(3–4), 223–264.
- Yadroff, M. (1991). *The syntactic properties of adjunction (scrambling in Russian)*. (Ms.)
- Yadroff, M. (1992). *Scrambling and relativized minimality*. (Ms.)
- Zemskaja, E. A. (1973). *Russkaja razgovornaja reč [Russian colloquial speech]*. Moscow: Nauka.
- Zemskaja, E. A., Kitajgorodskaja, M. V., & Shirjaev, E. N. (1981). *Russkaja razgovornaja reč: Obščie voprosy. Slovoobrazovanie. Sintaksis [Russian colloquial speech: General issues. Word formation. Syntax]*. Moscow: Nauka.

Samenvatting in het Nederlands

Deze dissertatie richt zich op twee afzonderlijke grammaticale problemen: opeenvolging van tijden (sequence of tense, SOT) en lange-afstandsverplaatsing. Russische en Engelse grammatica's verschillen ten opzichte van deze twee domeinen. Het doel van de dissertatie is om aan te tonen dat deze twee verschillen teruggebracht kunnen worden tot een enkel verschil in de opmaak van het C-T systeem in de twee talen. Dit verschil definieer ik in termen van de theorie van fasen (Phase Theory, Chomsky, 2001, 2005).

Ik richt mij op de volgende verschijnselen. Engels is een SOT-taal, hetgeen betekent dat verleden tijd in een deelzin ingebed onder verleden tijd in een matrix-zin normaalgesproken een simultane interpretatie krijgt, terwijl tegenwoordige tijd ingebed onder verleden tijd altijd gerelateerd is aan de tijd van uiting (speech time, cf. John said that Mary was pregnant; John said that Mary is pregnant). Het Russisch wordt tot de niet-SOT talen gerekend. In deze taal wordt verleden tijd ingebed onder verleden tijd normaalgesproken als temporele precedentie geencodeerd, terwijl tegenwoordige tijd ingebed onder verleden tijd gebruikt wordt om gelijktijdigheid uit te drukken zonder daarbij noodzakelijk naar de tijd te verwijzen waarop de taaluiting plaatsvindt.

Een tweede groep verschijnselen die ik analyseer betreft lange-afstandsverplaatsing. Het is bekend dat dergelijke verplaatsing zeer beperkt is in het Russisch. Zo is wh-verplaatsing uit een complementszin geïntroduceerd met de complementeerder *čto* "that" niet mogelijk. Dit feit is problematisch voor de huidige theorieën die de localiteit van verplaatsing beregelen, omdat dit type verplaatsing toegestaan zou moeten zijn. Het is interessant dat verplaatsing uit een subjunctieve complementszin (deze dragen de complementeerder *čtoby* "that") grammaticaal is in het Russisch. Extractie uit een indicatief complement is grammaticaal zodra de complementeerder weggelaten wordt. Tot dusver zijn deze feiten onverklaard gebleven.

Als theoretisch kader voor mijn onderzoek gebruik ik Phase Theory (Chomsky, 2001, 2005). Volgens deze theorie verloopt een zinsderivatie in

fases (phases). Phases zijn de eenheden van de syntactische structuur die naar het realisatie en interpretatiesysteem gestuurd worden zodra ze compleet zijn. Op deze manier worden ze ontoegankelijk voor verdere syntactische operaties. Alleen de periferie van de phase blijft “zichtbaar”. Het is belangrijk dat uitsluitend het hoofd van de phase een doel voor verplaatsing (of Agree - congruentie) kan zijn. Dit betekent dat alle verplaatsing of door moet gaan tot de rand van de phase, of er doorheen moet gaan in het geval van opeenvolgende cyclische beweging. Voor het Engels beargumenteert Chomsky dat CP en v*P phases zijn. Hij beweert bovendien dat hoewel TP phase-achtige eigenschappen heeft, deze niet op zichzelf een phase is: de TP verkrijgt deze eigenschappen van C en T kan alleen als afgeleide van C een doel voor verplaatsing zijn. Deze hechte relatie tussen C en T wordt geïllustreerd door het feit dat T niet zonder C kan voorkomen tenzij T defectief is, en dat de keuze voor de complementeerder strikt is verbonden met de finietheid van de deelzin.

Ik neem Chomsky's idee over dat C- en T-hoofden sterk aan elkaar gerefereerd zijn, en voor een concrete implementatie van dit idee richt ik mij tot Pesetsky en Torrego's (2001 et seq.) theorie van T-naar-C-verplaatsing. Kort gezegd beargumenteren Pesetsky en Torrego dat C altijd een oninterpreteerbaar T-feature draagt, dat normaliter zijn waarde van de locale T-hoofd krijgt. Een volgend belangrijke door mij overgenomen hypothese van Pesetsky en Torrego is het idee van congruentie (Agree) opgevat moet worden als het delen van features. Pesetsky en Torrego beargumenteren een kijk op de Agree-operatie waaronder de informatie over een niet ingevuld feature niet weggewist wordt zodra het feature een waarde krijgt, zodat er een relatie tussen de twee instanties van hetzelfde feature blijft bestaan. Het gevolg hiervan is dat Agree ook mogelijk is tussen twee niet ingevulde features: er wordt een relatie tussen hen gevormd, en beide (of alle) instanties van het feature kunnen dan een waarde krijgen als een van hen een Agree-relatie met een ingevulde instantie van hetzelfde feature aangaat. Volgens Pesetsky en Torrego is dit hoe C en T de waarde voor hun T-feature krijgen van v.

Ik gebruik het genoemde kader om een analyse van SOT in Engels en andere SOT-talen te kunnen leveren en om te kunnen verklaren waarom het Russisch zich anders gedraagt. Ik baseer mijn analyse op het idee dat tijdsmarkeringen (Tenses) referentiele entiteiten zijn, die bepaalde belangrijke eigenschappen met voornaamwoorden gemeen hebben. Erg belangrijk is het feit dat ze gebonden kunnen worden. Ik stel voor dat in het geval van SOT in talen zoals het Engels, de tijden in de ingebedde en inbeddende zinnen door binding met elkaar in een afhankelijkheidsrelatie staan. Deze relatie wordt geïmplementeerd als een syntactisch proces dat gebaseerd is op congruentie-relaties tussen de T-hoofden. Een dergelijk mechanisme is voorgesteld door Reuland (2001, 2005) om de binding van simplexe (SE) anaforen in talen zoals het Nederlands te analyseren.

Ik beargumenteer dat in het geval van de gelijktijdige lezing van ver-

leden tijd onder verleden tijd in het Engels, het ingebedde werkwoord de derivatie ingaat zonder waarde voor zijn temporele feature. Hierna wordt een Agree relatie gecreëerd tussen de C-, T- en v-hoofden in de ingebedde zinnen, aangezien C een niet ingevulde instantie van het T-feature draagt en zoekt naar een andere instantie van hetzelfde feature. Geen ingevulde instantie van het T-feature kan echter in de ingebedde clause gevonden worden. Als gevolg hiervan blijft de instantie van het T-feature oningevuld totdat de deelzin ingebed wordt in een hogere deelzin. Hierna wordt er een congruentie-relatie gemaakt tussen de ingebedde C en de hogere v. Als de hogere v een ingevulde instantie van het tense-feature draagt, zullen alle tense-feature-instanties in de ingebedde zin ingevuld raken. Op die manier krijgt het ingebedde werkwoord zijn tijds-waarde van het inbeddende werkwoord. Deze relatie wordt geïnterpreteerd als binding, hetgeen resulteert in een SOT-lezing.

Een dergelijke lezing kan niet in het Russisch verkregen worden. De reden hiervoor is dat de bindings-relatie niet gevormd kan worden. Ik stel voor het C-hoofd de zwakke link is in de reeks van Agree-relaties tussen de instanties van het tense-feature: in het Russisch draagt C niet een oningevuld T-feature. Dit betekent dat er via C geen congruentie-reeks gevormd kan worden tussen de T-hoofden in de ingebedde en inbeddende zin. Als het inbeddende werkwoord de derivatie ingaat met het tense-feature oningevuld, kan het dus geen waarde krijgen. Als gevolg zal de derivatie crashen.

Interessant genoeg kan soms een gelijktijdige lezing van verleden tijd onder verleden tijd verkregen worden in Russische complementszinnen. Ik laat met verscheidene voorbeelden zien dat deze lezing veel vaker voorkomt dan wordt aangenomen. Toch beargumenteer ik dat de aard van deze lezing verschilt van die in het Engels: in het Russisch wordt de ingebedde verleden tijd de re geïnterpreteerd. Het kan worden aangetoond dat de ingebedde verleden tijd door sprekers altijd geïnterpreteerd wordt als “echte” verleden tijd, relatief aan de tijd van het spreken, niet als tegenwoordig met betrekking tot de matrix-tijd. Bijvoorbeeld, “de aarde is rond” kan in het Russisch niet gebruikt worden in de verleden tijd wanneer het wordt ingebed onder verleden tijd. In het Engels is dit wel mogelijk.

Het verschil tussen de Russische en Engelse SOT-feiten ligt dan in de aard van de C-hoofden van de twee talen. Ik beargumenteer dat de eigenschappen van C voor nog een andere verzameling feiten verantwoordelijk zijn: de restricties op lange-afstands-extractie in het Russisch. Deze eigenschappen kunnen in termen van Phase Theory worden samengevat. Zoals ik eerder zei stelt Chomsky (2001, 2005) voor dat alleen phases - CP en vP - doelen van beweging kunnen zijn. Hij stelt dat TP geen phase is; T is afhankelijk van C en krijgt zijn temporele en ϕ -features van C. Hoewel dit geldt voor het Engels, in het Russisch is T en niet C een phase hoofd en dus is T onafhankelijk van C. In het Engels wordt dan het tense-feature door T geërfd van C; in het Russisch is dit niet het geval en dus worden er

geen overblijfselen van T op C gevonden. Deze claim wordt ondersteund door het feit dat de Russische complementeerder *čtoby* “that + subjunctief partikel” niet het hoofd kan zijn van niet-finiete deelzinnen. Dit betekent dat de aanwezigheid van de complementeerder *čto* “that” geen finietheid afdwingt. In het Russisch is er echter nog steeds een hechte relatie tussen C en T; de phase edge is nog steeds Spec,CP en niet Spec,TP.

Ik wijk af van Chomsky (2001, 2005) af waar hij argumenteert dat alle verplaatsing naar de rand van CP vrij is en gedreven wordt door “edge features” van C. Ik stel voor dat wh-verplaatsing gedreven wordt door een niet-ingevuld wh-feature, zoals algemeen wordt aangenomen. In het geval van opeenvolgende cyclische wh-beweging in het Russisch wordt dit feature toegekend aan alle T's, die de phase heads zijn en dus de enig mogelijke doelen voor verplaatsing. Op die manier verplaatst, in het geval van lange-afstands extractie over *čto* de wh-frase naar een “intermediate landing site” aan de rand van TP. Echter, aangezien Spec, TP niet de rand van de phase is, zal de wh-frase daar gefixeerd zijn en niet door kunnen verplaatsen naar zijn eindbestemming in de matrix-zin. Dit leidt tot een derivatieve crash.

In tegenstelling tot dit, zal T in *čtoby*-deelzinnen naar C verplaatst worden, zoals men kan zien aan het feit dat het partikel *by* (hetgeen deel is van de werkwoordsvorm) samengevoegd wordt met de complementeerde *čto*. En dus bereikt het door T aangetrokken wh-element Spec, CP, hetgeen zich aan de rand van de phase bevindt en zichtbaar is voor volgende verplaatsingsoperaties. Ik stel voor dat wanneer de complementeerder *čto* weggelaten is, extractie grammaticaal wordt doordat de phase-rand door het systeem in fonetische termen opgevat wordt. (Een dergelijk idee wordt ook geopperd in Mulders, 2002, waar voorgesteld wordt dat de parser soms alleen een functionele categorie in betrekking neemt wanneer deze overt gerealiseerd is.)

Curriculum vitae

Olga Khomitsevich was born on the 24th of May 1981 in St. Petersburg, Russia (Leningrad, USSR at that time). In 1998, she graduated (cum laude) from St. Petersburg Classical School 610. She entered the Department of Theoretical Linguistics at St. Petersburg State University in 1998 and received her MA degree (cum laude) from St. Petersburg State University in 2003, majoring in general linguistics and specializing in psycholinguistics.

In 2003, Olga Khomitsevich started her PhD research at the Department of Theoretical Linguistics at St. Petersburg State University. In 2004, she enrolled in the joint PhD program organized by Utrecht Institute of Linguistics OTS and St. Petersburg State University. This dissertation is the result of the work she carried out as a part of this program.