Martina Lindseth

Null-Subject Properties of Slavic Languages

With Special Reference to Russian, Czech and Sorbian

Slavistische Beiträge

Begründet von
Alois Schmaus

Herausgegeben von
Peter Rehder

BAND 361

Martina Lindseth

Null-Subject Properties of Slavic Languages

With Special Reference to Russian, Czech and Sorbian

PVA

98.

2677

ACKNOWLEDGEMENTS

This work originates from my Ph.D. thesis which I defended at Indiana University, Bloomington, U.S.A. in 1996. I would like to thank all members of my dissertation committee, especially Dr. Steven Franks, for guiding me through my graduate studies at IU and the preparation of my dissertation. I would also like to express my graditute to the scholars at the Serbski institut in Budyšin. Special thanks go to Dr. Irene Šěrak for helping me find native speakers for my fieldwork, and for reading portions of the manuscript, and to Dr. Helmut Faßke for allowing me access to some of his recent data. I would also like to thank Erik for all his support and patience while I was working on the dissertation, and Pat for proofing the book manuscript.

This research was supported in part by an IU International Programs Pre-Dissertation Fellowship, an IU College of Arts and Sciences Summer Research Fellowship, a Wabash College Faculty Development Grant and an American Council of Learned Societies Dissertation Fellowship. All financial help was greatly appreciated and allowed me to expand the project far beyond my original goals.

This book is dedicated to my father, Dr. Siegfried Profe, who always rejoiced in my successes. He would have celebrated the completion of this work.
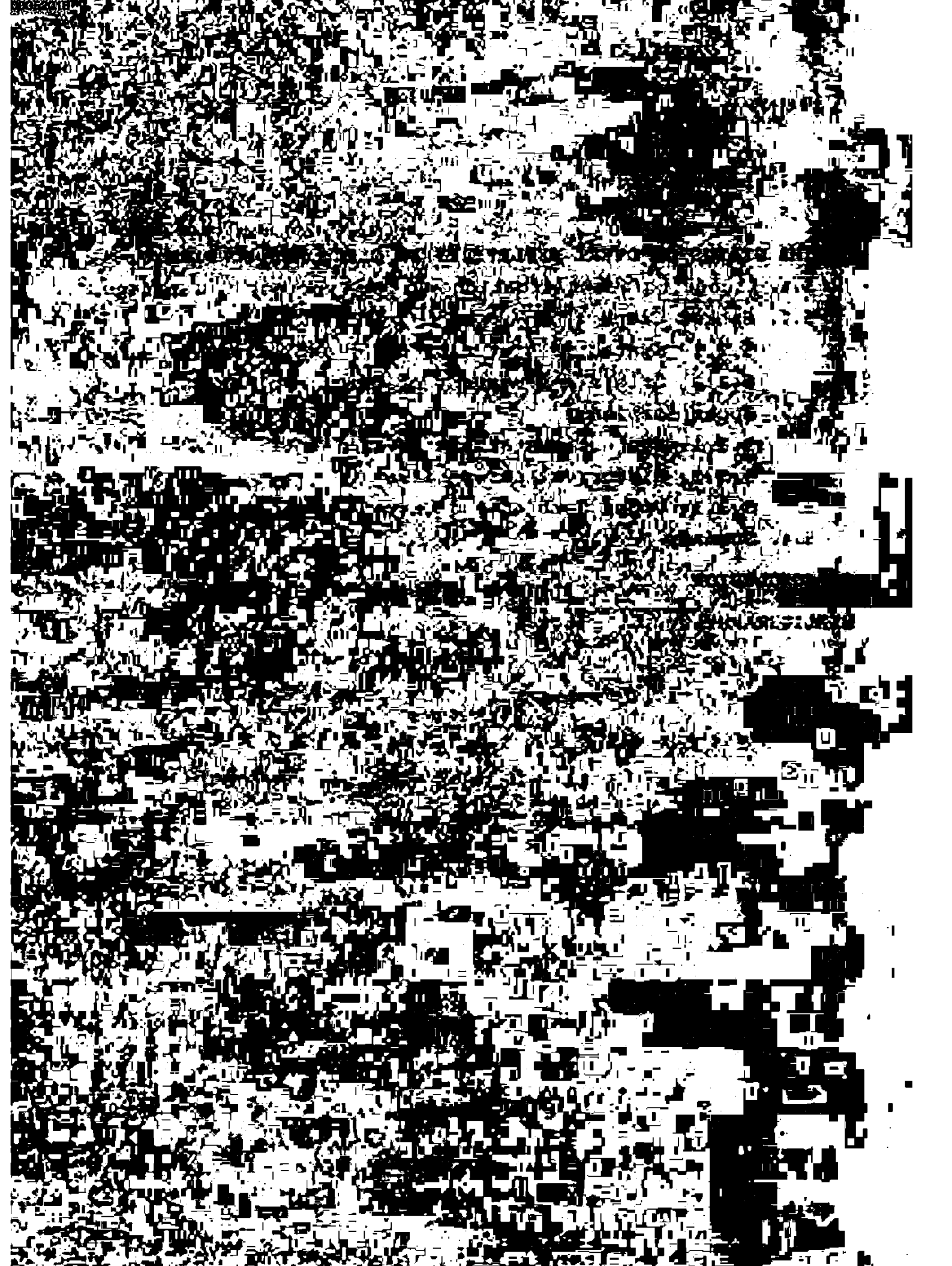
Eau Claire, Wisconsin, summer 1998

-VI-

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

1. INTRODUCTION

1.1. The issue of null subjects

It is readily apparent that languages vary according to whether an unstressed pronominal subject of a finite clause is lexicalized (English, French) or not (Spanish, Italian). This study derives from and contributes to the large corpus of recent research by generative grammarians which has sought to explain this crosslinguistic variation in the occurrence of null subjects.

In a broader sense this work is part of an even larger tradition since generative syntacticians were certainly not the first to concern themselves with the zero-problem in linguistics. Indeed, the concept of zero-elements (also called null elements or phonologically empty elements) has been previously invoked by many linguists belonging to different schools of linguistics, including those of the French linguist de Saussure and his students, Jakobson and the Prague functionalists, and Bloomfield and the American Structuralists. As a result Meier (1961) was able to devote an entire monograph to the "Zéro-Problem in der Linguistik", in which he included a survey of "Zéro-Spezial-Arbeiten" and a survey of the history of this topic in linguistics.

For generative linguists, null subjects have proven to be a fruitful topic of debate ever since Perlmutter's *Deep and surface structure constraints in syntax* (1971). Since one of the central concerns of the "principles and parameters" model

of government and binding (GB) theory has been to determine positions in which overt and phonologically empty manifestations of different NP types can occur, defining the precise nature of the crosslinguistic variation in the availability of empty subjects has been of special interest to GB, generating numerous articles and several books, such as Bouchard (1984), Roberge (1990), and Jaeggli and Safir (1989).

Traditionally, Italian has been cited as an example of the canonical null-subject or *pro*-drop languages which are clearly identifiable as different from canonical overt subject languages, such as English, as illustrated in (1).

(1) parlo   I speak

    parli   you speak

    parla   he speaks

    parliamo   we speak

    parlate   you speak

    parlano   they speak

Since the Italian inflectional paradigm uniquely distinguishes all person/number combinations, it is intuitively clear, independent from any specific grammatical framework, why Italian can afford to do without overt subject pronouns: The pronominal features (person, number) of a "missing" subject can be recovered unambiguously due to the rich inflectional agreement on the verb.

However, with the eventual spread of theoretically in-formed generative studies to more and more languages, it be-came apparent that the situation is far more complex than expected. It is now clear that the relationship between rich agreement and null subjects is less direct than originally assumed. On the one hand, there are languages like German, which seem to have what we would call rich agreement, but fail to allow referential null subjects, while on the other hand, there are languages like Japanese and Chinese, which are quite capable of omitting subject pronouns, despite their lack of agreement morphology.

Further research revealed that some languages (such as German, Icelandic and Russian) can show mixed behavior with respect to the *pro*-drop parameter. In German, for instance, referential subjects and quasi-argumental[1] subjects are not omitted, whereas truly expletive subjects are.

(2) ich/*Ø liebe Tiere
'(I) love animals'

(3) draußen regnet es/*Ø schon wieder
'outside it is raining again'

(4) ich bin froh, daß Ø/*es getanzt wird
'I am happy that there is dancing going on'

---

[1] As pointed out by Chomsky (1981: 325), quasi-argu-ments have some properties of true arguments (e.g. control of PRO) and some properties of non-arguments (cannot be replaced by referential NPs) and thus behave differently than truly expletive subjects.

Thus there is in fact no clearcut division between null-subject and non-null-subject languages. Despite the longstanding intuition that there is a relationship between richness of agreement and the availability of null subjects in a specific grammar, a definitive and completely accurate characterization of the notion of inflectional richness has yet to be proposed.

1.2. Slavic languages and universal grammar

The theoretical parts of this work are conceived within the syntactic framework of GB theory, initiated by Chomsky (1981, 1982) as a development of earlier versions of generative grammar and further refined during the past decade. The crucial hypothesis adopted by Chomskian syntacticians is that a large part of a speaker's knowledge about language is innate. All humans are endowed with the same universal grammar (UG), which consists of a set of absolute universals and principles, which do not vary from one language to the next, and a range of parameterized principles whose actual values have to be fixed during childhood language acquisition upon exposure to language. Thus, while the language-specific properties themselves are not fully determined by UG, the available options are. The grammar of a specific language boils down to a particular set of values for the parameters. This theory, for apparent reasons, is often referred to in the relevant literature as the "principles and parameters" theory and we shall use this term synonymously with "government and binding".

Until fairly recently the Slavic languages received relatively little attention within GB theory. This is actually quite surprising, considering that these 14 closely **related** languages[2] show interesting syntactic **differences** which can provide Chomskian syntacticians with an ideal pool of data to test their theories of universal principles and potential parametric variation. Certainly, from the perspective of null-subject phenomena, which are the focus of this work, the Slavic languages are quite diverse.

The identification of particular problems in the Slavic languages may thus indicate modifications which must be made to current theories in general. Since it is a basic tenet of UG that any analysis proposed for some language(s) should be able to explain data of other languages and accommodate cross-linguistic variation, it follows that the investigation of data from "overlooked" languages is often crucial in proving or disproving the general validity of theoretical assumptions.

In this context perhaps the most neglected of the modern Slavic languages is Sorbian, also referred to as Lusatian and Wendish, which belongs to the West Slavic language family. Since it is one of the most complex Slavic languages with regard to null subjects, it is a major focus of this present

---

[2]   East Slavic: Russian, Ukrainian, Belorussian
West Slavic: Polish, Kashubian, Upper Sorbian,
  Lower Sorbian, Czech, Slovak
South Slavic: Serbian, Croatian, Slovene,
  Macedonian, Bulgarian
There is some disagreement among linguists whether Kashubian and Polish, and Serbian and Croatian are separate languages.

work. Sorbian is generally accepted to have split into two different literary languages, identified as Upper and Lower Sorbian, which are spoken by a very small minority in and around the cities of Bautzen and Cottbus in Eastern Germany. The total assimilation of this Slavic society into the vastly greater German one which surrounds it also plays a role in this study, culturally as well as linguistically.

The rapid decrease in the number of native speakers of Sorbian in recent decades (about 22,500 remain) has been accelerated by the economic and social disruption brought about by the collapse of East Germany in 1989. It is hoped that this current research will help to slow this decline by increasing awareness, and contributing to the (re)introduction of the Sorbian languages into academic discussion. Ironically, the same political changes which threaten the survival of the Sorbian language and culture have also improved access to native speakers of Sorbian by Western linguists.

1.3. Organization

This work is divided into four main chapters, the first of which, Chapter 2, sets out to establish a theoretical background. The **licensing** and **identification** approach to null subjects is summarized and a set of three diagnostics introduced which can be used to decide whether or not a language has all the properties of a canonical null-subject language.

In Chapter 3, these diagnostics are applied to Slavic in general, identifying a split between East Slavic on the one hand, and West and South Slavic on the other. This is followed by a section which discusses why the Slavic languages are divided along these lines.

Chapters 4 and 5 are devoted to Upper Sorbian and Czech. Both chapters present and discuss linguistic data which seem to challenge the licensing/identification framework. This begins with Upper Sorbian which provides a telling counter-example to the long accepted intuition that the explicitness of verbal morphology and occurrence of null- pronominal subjects are directly related. A broad examination of contemporary and early texts reveals a surprisingly high level of overt pronoun use, which makes Sorbian an exception among the West Slavic languages. Chapter 5 focuses on the comparison and analysis of the occurrence of overt expletive subjects in Czech and Upper Sorbian. Czech generally omits unstressed pronominal subjects. This is expected, since Czech verb forms uniquely distinguish the different person/number combinations, very much like in the Italian example above. However, overt non-referential pronouns occur frequently in spoken Czech, which has generally been assumed impossible for languages which omit referential subjects. The analysis of Czech expletive subjects is followed by a discussion of overt non-referential pronouns in Upper Sorbian.

2. NULL SUBJECTS IN THE GOVERNMENT BINDING FRAMEWORK

2.1. Theoretical background

GB theory describes knowledge of language as a highly struc-
tured theory of UG which is composed of an interrelated set of
six modular subtheories. These are enumerated by Chomsky
(1981: 5) as bounding theory, government theory, theta-theory,
binding theory, Case theory and control theory. These sub-
systems, in turn, are made up of general principles and para-
meters whose potential value settings vary within clearly
defined limits from one language to another. GB is also modu-
lar in the sense that it assumes four different levels of
representation, known as D(eep)-structure, S(urface)-struc-
ture, phonetic form (PF) and logical form (LF). These differ-
ent levels of representation are related to each other by a
transformational component which contains the rule Move-α,
which, in effect, allows anything to move anywhere. Various
general principles regulate the mapping from one level to
another by determining which movements are well-formed and
which are not.[1] One important constraint on the relation be-
tween syntactic representations is the structure-preserving

---

[1]   There are central structural concepts which are not
confined to any particular subtheory, most notably the no-
tions of "c-command" and "government". Here, we assume the
following definitions (simplified from Chomsky, 1986):
C-command (extended to include m-command):
Node A c-commands node B iff A does not dominate B and every
X that dominates A also dominates B.
Government:
A governs B iff A c-commands B and no barrier intervenes
between A and B.

principle, derived from the work of Emonds (1976), according to which movement operations must not destroy existing structures. This means that those structures established at D-structure must be preserved at S-structure. It follows that when a given element is moved from its extraction-site to its landing-site it cannot simply disappear from its original site "without a trace". Hence, whenever an element is moved, a coindexed, empty element called a trace remains in the moved element's original position.

Theory-internal reasons force this syntactic framework to assume the existence of non-overt, but structurally present categories. Two principles, the theta-criterion and the projection principle, which define how the properties of lexical entries are projected onto syntax, are especially crucial in this respect. The theta-criterion specifies that each argument bears one and only one theta-role and that each theta-role is assigned to one and only one argument. According to the projection principle, the specifications for each lexical item given in its entry in the lexicon have to be met at all levels of syntactic representation. In other words, the one-to-one pairing between the thematic roles and arguments of a verb's predicate-argument structure required by the theta-criterion has to remain consistent on all levels of derivation (D-structure, S-structure, and LF). The projection principle and the theta-criterion determine whether an NP has to be present in a syntactic representation. However, a structurally

required NP can be manifested either as phonologically realized (overt) or non-realized (null). The latter is called an empty category. The various sub-theories of GB interact to determine whether the NP is lexical or not, and what type of NP appears in any given position. Certain empty categories can only appear in certain structural positions, which is reflected in the formulation of licensing requirements for empty categories. However, the exact formulation of such licensing conditions is far from being uniformly agreed upon. Furthermore, the content of those "missing" elements has to be identifiable. The position of traces, for instance, is formally licensed by the empty category principle,[2] while their content is determined by their antecedent. An interesting issue in this context is whether formal **licensing** of the occurrence and the **identification** of the content of an empty category are achieved by one and the same or two distinct operations. We shall return to this question in section 2.3.

The typology of empty categories as developed by Chomsky (1982: 78-89) mirrors that of overt categories. The module of the grammar regulating the interpretation of NPs is the bind-

---

[2]   Traces must be governed in a special way in order to occur in certain positions: They must be **properly** governed.
Empty category principle:
Traces must be properly governed. *A* properly governs *B* iff *A* theta-governs *B* or *A* antecedent-governs *B*. (Chomsky, 1986: 17).
A head *A* theta-governs *B* if it governs and theta-marks *B*. A maximal projection *A* antecedent-governs *B* if *A* and *B* are coindexed and *A* governs *B*.

ing theory. Chomsky proposes that the three overt NP-types, anaphor, pronoun and R-expression, should be interpreted as composed of two features: [±anaphoric] and [±pronominal].[3] Depending on its feature matrices, the NP is subject to different principles of the binding theory. An NP-type which is subject to principle A of the binding theory has the feature [+anaphoric] and consequently has to be bound in its governing category. A category which is subject to principle B of the binding theory has the feature [+pronominal] and therefore has to be free in its governing category. Chomsky further assumes that the typology of empty categories can also be derived from the possible combinations of [±anaphoric] and [±pronominal].

The inventory of overt and empty NPs is illustrated in Table 1.

Table 1: Classification of NPs

Anaphoric	Pronominal	Overt NPs	Empty NPs
+	-	anaphors	NP-traces
-	+	pronouns	*pro*
+	+	excluded	PRO
-	-	R-expression	Wh-traces (variables)

---

[3]   This is analogous to the representation of syntactic categories such as N, V, P and A by feature matrices rather than primitives.

According to Table 1, anaphors and NP-traces are subject to principle A of the binding theory and must be bound within their governing category. Pronouns and *pro* have to be free within their governing category, and referring expressions and variables are subject to principle C, that is they must not be bound anywhere. The features proposed for the NP-types are binary, and, in the ideal case, we expect all four combinations to occur. The inventory of empty categories supports this expectation. However, there are only three overt NP-types. The occurrence of an overt NP-type with the feature composition [+anaphoric,+pronominal] is ruled out by independent principles of UG.[4]

The different overt and null elements are exemplified in the following sentences. The Slavic examples are from Russian unless otherwise indicated:

---

[4]   An element which is [+anaphor] must be bound within its governing category, in keeping with principle A of the binding theory. An element which is [+pronominal] must be free in its governing category in keeping with principle B of the binding theory. NPs with the features [+anaphoric, +pronominal] are thus subject to contradictory requirements. The only way for an element to fulfill both principle A and B at the same time is to lack a governing category. If there is no governing category, then neither principle A nor principle B apply, because the domain in which the element has to be both free and bound is not defined. This can be fulfilled under the condition that the NP lacks a governor. However, an overt NP that lacks a governor cannot be case-marked, under the assumption that case marking requires government, and hence will be ruled out by virtue of the case filter, according to which all overt NPs have to be case marked. As an empty category, PRO is not subject to the case filter and the absence of a governor is not problematic. Furthermore, due to its feature composition PRO is only admitted in ungoverned positions. This proposition is referred to in the literature as the "PRO theorem".

-13-

(1) [+anaphoric,-pronominal]

Overt: anaphor

Ivan$_i$ ljubit tol'ko sebja$_i$

'John$_i$ loves only **himself**$_i$'

Empty: NP-trace

dom$_i$ byl postroen t$_i$

'the house$_i$ was built t$_i$'

(2) [-anaphoric,+pronominal]

Overt: pronoun

Ivan$_i$ nenavidet ego$_{y*i}$

'John$_i$ hates **him**$_{y*i}$'

Empty: *pro*

pro miluju Mariu   [CZ]

'(I) love Mary'

(3) [+anaphoric,+pronominal]

Overt: excluded

Empty: PRO

on xočet **PRO** ženit'sja na Marii

'he wants [**PRO** to marry Mary]'

(4) [-anaphoric,-pronominal]

Overt: R-expression

on$_i$ dumal, čto Ivan$_{y*i}$ pobedil v zabege

'he$_i$ thought that **John**$_{y*i}$ won the race'

Empty: variable

kogo$_i$ vy priglasili t$_i$

'whom$_i$ did you invite t$_i$'

A number of issues connected with the types of empty categories described above remain controversial, although the study of empty categories with respect to their type, distribution and contents has become a very popular topic within GB.

2.2. Are null subjects more than just nothing?

Let us review some of the empirical and theory-internal reasons which support the assumption that empty subjects of both tensed and infinitival clauses are real in the sense that they are structurally present in a representation. Following Chomsky (1986) I assume that X-bar theory, including the endocentricity requirement for phrases, applies to both lexical and non-lexical (functional) categories. The clausal category traditionally labeled S is reinterpreted as the maximal projection of I, where I=INFL, and the category S' is reinterpreted as the maximal projection of C, where C=complementizer.

(5) S'

Comp   S

    NP INFL   VP

(6)   CP

⇒   Spec   C'

          C   IP

              NP   I'

                  I   VP
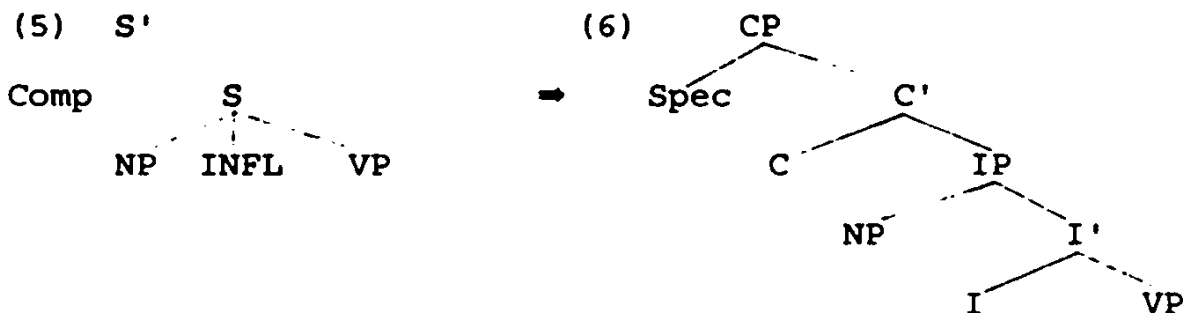
Unless otherwise stated, I will assume that structural sub-
jects are base-generated in and occupy the specifier position
of IP [NP,IP].[5]

In the following I argue against (7a) and in favor of
structure (7b) for all sentences, even if the subject is not
phonologically realized:

(7) a.   b.

```
            IP                                 IP
            '                                /  '  \
            I'                             NP      I'
         /     \                                /     \
       I         VP                          I          VP
```
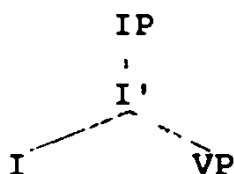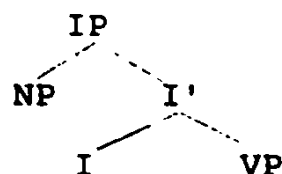
In the sentences below it is obvious that even in the absence
of an overt subject its presence is felt implicitly and its
content can (and must) be recovered to ensure a complete
statement. (Here and in later examples I shall use e to indi-
cate the presence of a generic empty category)

(8) e biju ho   [Cz]
     '(I) beat him'

(9) on xočet [e bit' Ivana]   [Ru]
     'he wants to hit Ivan'

The verb 'to hit' takes two arguments: an external argument
which receives the theta-role of agent, and an internal argu-

_____

[5]   An alternative proposal which has been gaining
ground in the recent literature is the so-called VP-internal
subject hypothesis. Sportiche (1988) proposes that all sub-
jects are base-generated in [Spec,VP] position, so that NPs
in [Spec,IP] are derived subjects. I shall return to this
hypothesis in subsequent chapters.

ment which is assigned the role of patient. It follows from the theta-criterion in conjunction with the projection principle that a sentence with a verb which assigns an external theta-role must have a structural subject at any level of the representation.

Within GB, the missing subjects in (8) and (9) represent two fundamentally different types of empty categories, although they have in common the fact that they are base-generated, as opposed to the [-pronominal] empty categories, which represent traces of movement. The non-overt subject of an infinitive is PRO, while missing subjects of tensed clauses are *pro*. In contrast to PRO, which must be ungoverned according to the PRO-theorem, *pro* is found in governed position since it alternates with overt NPs. It is generally assumed that nominative case is assigned to an NP in the specifier position of IP through government by a finite [+tense,+AGR] INFL. A non-finite [-tense,-AGR] INFL on the other hand, is not a governing head, leaving [Spec,IP] ungoverned so that PRO occurs in that position.

(10) a.   b.

```
        IP                              IP
       /  \                            /  \
     pro    I'                      PRO      I'
            / \                            /  \
           I   VP                         I    VP

        +tense                          -tense
        +AGR                            -AGR
```

The central concern for this study is the different manifesta-
tions of subjects of finite clauses (as *pro* or as lexicalized
pronoun) in Slavic languages, with a special focus on Russian,
Czech and Upper Sorbian.

More evidence supporting the structural presence of a
subject position can be drawn from the modules of binding and
control theories. Consider the following Czech sentences and
the binding facts they demonstrate:

(11) e miluje sebe   '(he$_i$) loves himself$_{i/j}$'

(12) e miluje ho   '(he$_i$) loves him$_{*i/j}$'

(13) e miluje Petra   '(he$_i$) loves Peter$_{*i/j}$'

These facts follow immediately from principles A, B and
C of the binding theory, if we assume the presence of an (emp-
ty) element in [NP,IP] position.

The following examples show that the theta-criterion
alone cannot justify the existence of all empty subject NPs:

(14) e temneet   [Ru]
      '(it) is getting dark'

(15) a. e prší   b. e zdá se, že   [Cz]
          '(it) is raining'   '(it) seems that'

The verbs in (14) and (15) do not assign any obvious external
theta-role to the NP in subject position.[6] However, there is

---

[6]   According to Chomsky (1981: 325), subjects of so-
called weather-verbs have a "quasi-argumental" status. For
some languages such a distinction of non-referential subjects

another (admittedly somewhat dubious) principle within GB, known as the extended projection principle (EPP), which stipulates that all sentences must have a subject, regardless of their argument structure. Here I follow the traditional GB analysis according to which in the sentences above a null-expletive subject occupies the structural subject position. Similarly, in overt subject languages, such as English, the EPP accounts for the presence of the expletive *it* or *there* in subject position.

(16) *e/it is nice that you came

However, some researchers, including Babby (1989), reject the EPP and instead postulate the existence of truly subjectless sentences, i.e. sentences without a structural subject position. Of course, the existence of null expletives raises obvious empirical problems. Why should they exist, (after all they are invisible **and** have no meaning), and how do we know and learn that they exist? One might simply want to claim that the obligatory structural presence of a subject position is part of UG. Chomsky (1982: 10) does exactly that when he writes about the EPP:

---

may be crucial. In German, for instance, quasi-argumental subjects, as in (ii), and non-argumental subjects, as in (i), behave differently: The latter must remain lexically unexpressed, while the former must be overt.

(i) ich gehe nicht raus, weil e/*es dort getanzt wird
    'I don't go outside because there is dancing going on'

(ii) ich gehe nicht raus, weil es/*e regnet
    'I don't go outside because it is raining'

We may think of it as a general principle governing
D-structures, hence also governing structures de-
rived from them. There are good reasons, which I
will not elaborate here to assume that the same is
true of languages that appear to allow a "missing
subject" (e.g. Romance languages other than French).

It is interesting that other theories also postulate the

existence of such covert expletives due to theory internal

requirements, such as Relational Grammar's "Final-One Law".

Be this as it may, while currently we cannot really prove that

these empty elements are structurally present, it is at least

as impossible to prove that they do not exist, due to their

special "no meaning" and "no lexical material" property.

2.3. Development of the *pro*-drop parameter

Languages that leave a pronominal subject unexpressed tradi-

tionally are called *pro*-drop languages, since they seem to

"drop" the subject pronoun. This term stems from earlier

analyses and reveals that the phenomenon of null subjects was

originally interpreted as a deletion rule (Perlmutter 1971).[7]

Thus, from the perspective of the current syntactic framework

the term "*pro*-drop" is obsolete, since the deletion of a sub-

ject in the course of a derivation would contradict several

---

[7]   In earlier accounts it was claimed that the *pro*-
drop phenomenon correlated to a host of other syntactic prop-
erties, such as free subject inversion and apparent viola-
tions of the that-trace filter, suggesting that a single
grammatical property might account for a whole complex of
contrasts between null-subject and overt-subject languages
(cf. Perlmutter 1971; Taraldsen 1978; Jaeggli 1982; Rizzi
1982). This idea was later challenged by Safir (1985), among
others.

theoretical assumptions, most notably the theta-criterion and structure-preserving principle. However, in keeping with the current literature on the topic, we shall maintain the label *pro*-drop parameter here. The development of the *pro*-drop parameter over the years reflects one of the most important tendencies of GB theory in general: a movement away from specific rules and towards general principles.

Since Taraldsen (1978), the *pro*-drop parameter has been associated with the phenomenon of "rich inflection" which can determine the grammatical features of a missing subject. He analyzed null subjects as base generated anaphoric NPs, rather than as the result of "pronoun deletion". According to him, empty subjects are null anaphors which do not violate Chomsky's nominative island constraint[8] on nominative anaphors due to the fact that the null-subject NP and the finite verb are coindexed. Thus, in languages with rich agreement the subject-verb agreement can bind a null-subject NP, thereby rendering the NIC inoperative. Taraldsen's analysis can be rephrased within a theory which now contains the ECP instead of the NIC. A subject pronoun can be omitted if AGR is rich enough to properly govern the subject position.

As far as determining the type of the empty category in the subject position of tensed clauses, the investigation of null subjects predates the discovery of *pro*. Chomsky (1981)

---

[8] The NIC (replaced by the ECP) stipulated that nominative anaphors have to be bound within their clause.

still assumes that there are only two types of empty catego-
ries, PRO and traces, and that the missing subject of a tensed
clause is of the PRO type. As we saw earlier, PRO is charac-
terized within the GB framework as a pronominal anaphor, and
under that assumption it has to be ungoverned. Chomsky (1981)
then has to stipulate that in null-subject languages the sub-
ject of tensed clauses may be ungoverned at S-structure due to
movement of INFL into VP in the syntax. This analysis contra-
dicts those by Taraldsen (1978) and Rizzi (1982),[9] where the
empty subject is said to be licensed precisely under govern-
ment by INFL. It is not until his 1982 monograph that Chomsky
posits the existence of the empty pronominal category *pro*, in
the process of developing the full typology of empty elements.

Most approaches since Taraldsen have identified the cru-
cial property of null-subject languages in the inflectional
element, usually relying on the notion of richness in one way
or another. One of the earliest and most influential GB ac-
counts of the null-subject parameter is given by Rizzi (1982).
Perhaps the most important aspect of Rizzi's paper is his
argument that what had been considered a monolithic parameter
of UG should really be split into two autonomous yet connected
subparameters. His argument is based on the different distri-
butional restrictions of referential and "dummy" empty sub-
jects in Italian. The first subparameter specifies whether a

---

[9]   Rizzi (1982) was already available as a manuscript
in 1980 and Chomsky (1981) refers to it in several places.

language licenses null subjects (INFL [+pronominal]). This property is **not** directly related to the morphological richness of the language. The second subparameter, however, is direct-ly related to morphological richness and determines whether INFL can recover the referential content of an empty subject (INFL [+referential]. These two parameters predict four theo-retically possible language types (Rizzi 1982: 143):

1	2	3	4
+pron	+pron	-pron	-pron
+ref	-ref	+ref	-pron

Language type 1 is represented by Italian, a canonical *pro-drop* language. Type 4 is a canonical non-null-subject lan-guage like English. According to Rizzi, combination 3 is excluded for "intrinsic reasons". A pronominal INFL can be referential only if its grammatical features of agreement are structured beyond a certain richness threshold. In other words, an inflection cannot be [+ref] if it is not [+pron]. Type 2 represents a language type which has phonetically null dummy subjects but overt referential subjects. Rizzi (1982) points out that this language type seems to exist, however, he does not go into much detail.[10]

---

[10] Rizzi (1982: 143) cites the northern Italian dia-lect of Padua as a potential example for this parameter com-bination. Later research provided more evidence for this "mixed" language type (e.g. Russian, German, Icelandic).

For any analysis of empty elements, null subjects includ-
ed, there are two questions which need to be answered: What
formally licenses them? How is their content (the person-
number features of the missing subject) determined? Early
approaches make no clear distinction between licensing and
identification. Even for Rizzi (1982), who in principle argues
in favor of a split of the *pro*-drop parameter into licensing
and identification, it is INFL which licenses the null subject
of a tensed clause, and it is the spelling out of the pronomi-
nal features of INFL which allows identification of the miss-
ing subject.

Early work on the topic of null subjects concentrated
almost exclusively on rich agreement languages, and thus it
was supposed to be rich agreement alone which made *pro* possi-
ble. It was not until other languages were included in the
research that it became apparent that the situation was far
more complex.

Rizzi (1986) criticizes the standard approach under which
pro is restricted to occur only in the context of rich agree-
ment and which unifies the formal licensing of *pro* and the
identification of its content into one and the same operation
via government by a strong Agr. He points out that in the
broader context of null elements such an assumption is sur-
prising, since for traces and PRO the licensing conditions and
the recovery operations are kept separate. Formal licensing
appears to involve some kind of government relation, while the

recovery procedure involves some kind of binding relation. Based on the occurrence of null objects (*pro*) in Italian despite the lack of object agreement, Rizzi (1986) argues in favor of divorcing the licensing conditions for *pro* from the inflectional system. He proposes what has become known as the *pro*-module, made up of two main independent principles (Rizzi 1986: 521,524):

(17) *Pro*-module:

A: *pro* is licensed if case-marked by X_y^0
B: *pro* has the grammatical specification of the features on X coindexed with it

The class of licensing heads can vary from language to language and seems in a sense arbitrary. In Italian, for instance, INFL is a licensing head for *pro* in subject position, and V is a licensing head for *pro* in object position.[11]

Jaeggli and Safir (1989) maintain the sharp distinction between licensing and identification, however, they present a novel treatment of the two processes. Chinese and Japanese, both null-subject languages which completely lack any sort of agreement, show clearly that the licensing criterion cannot be exclusively associated with rich morphology. Jaeggli and Safir's generalization is that null subjects seem to be licensed in languages with either very rich agreement or no

---

[11] Rizzi's (1986) study implies that *pro* is not restricted to subject position of tensed clauses. However, null subjects remain the main focus of this study.

agreement at all, and they formulate morphological uniformity as a licensing criterion.[12] An inflectional paradigm of a language is morphologically uniform if it has either only underived inflectional forms or only derived inflectional forms. If some forms of a paradigm are divisible into stem+affix and other forms of the same paradigm are bare stems, then the paradigm is mixed, and no null subjects are licensed. Under this model, the licensing conditions do not provide a mechanism to recover the referential value of a null pronoun. A separate identification mechanism is necessary to determine the content of a thematic NP. According to Jaeggli and Safir, identification can come about in a couple of different ways. However, they are fairly vague on formulating the precise conditions for identification. Identification can happen locally by agreement if the category containing AGR case-governs the empty category.[13] Thus, in languages like

---

[12]   The discovery that the phenomenon of "*pro*-drop" can occur when there is a rich agreement element or when there is no agreement at all actually goes back to Huang (1984).

[13]   It has been argued that in V-2 languages, such as German or Icelandic, the nominative case is assigned by Tense, which is located in COMP (C of CP), while AGR is located in INFL (I of IP). Under the requirement that identification by AGR can only happen if the category containing AGR is case-governing the empty subject, we now have an account for the lack of thematic null subjects in German and Icelandic. Both languages have morphologically uniform paradigms, and hence license null subjects, however, identification is blocked due to the distribution of Tense and AGR in separate nodes. More support for this approach comes from West Flemish (cf. Bennis and Haegemann (1984), which has complementizers which are inflected for person and number and indeed allows null thematic subjects. This is, then, an instance in which both Tense and Agr are part of the same governing node, and identification can take place.

Spanish or Italian, the agreement affixes with the relevant *phi*-features act as identifiers. For languages which uniformly lack agreement, such as Chinese and Japanese, Jaeggli and Safir essentially follow Huang's (1984) analysis. In these languages, a null subject can be identified either by a c-commanding nominal in the matrix clause or by a (null) topic.[14]

Although it is not clear *per se* why something like the morphological uniformity criterion should be operative in

---

[14]   Only in the first instance should we really speak of *pro*. Consequently, the occurrence of genuine null subjects in Chinese and Japanese is in fact quite restricted. Consider the following Chinese examples (Huang 1984):

Speaker A: Zhangsan kanjian Lisi le ma?
             'did Zhangsan see Lisi'

Speaker B: (i)   e kanjian ta le
                  '(he) saw him'

           (ii) Zhangsan xiwang [e keyi kanjian Lisi
                'Zh. hopes (he) can see Lisi'

In Huang's analysis, the empty element in (i) can only be a variable bound to a zero topic. In "rich" agreement languages which license *pro* the empty element in the same position could be identified by Agr. This type of identification is unavailable in Chinese. The empty category in (ii) may have dual status. Its reference can be determined by an "outside" discourse topic (variable reading) distinct from *Zhangsan* or it can be fixed via coindexation with the NP *Zhangsan* in the matrix clause (pronominal reading).

One might object that even in a *pro*-drop language the reference of an empty third-person subject, such as in (i), seems to depend on the discourse. The difference, however, is the following: The interpretation of 3rd person *pro* could be understood as a two-step process. First, the pronominal features under AGR determine the pronominal content (person, number, gender) of the empty pronoun. This is what I refer to as identification of the content of *pro*. Once that is done, the identified *pro* behaves like a "normal" pronoun, being interpreted according to principle B of the binding theory. During that part of the interpretation operation a pronoun may be freely coindexed with NPs outside their governing domain.

natural language (cf. Speas (1994) below for a possible inter-

pretation), the licensing and identification theory developed

by Jaeggli and Safir fits well into the general context of

empty categories. Similar to Rizzi's (1986) *pro*-module, it

keeps licensing and recovery of content separate. Licensing

is a purely formal criterion, while identification is a more

contentful criterion, referring to the availability of some

mechanism to retrieve the essential grammatical information

left phonologically unexpressed in order to provide the empty

category with some content. For traces, this division of

labor is exemplified by the ECP as the licensing condition,

and antecedent government as the identification operation.

The empty category PRO is limited to occur in ungoverned con-

texts, and derives its content from an antecedent which is

determined through control theory.

Originally it was assumed that there were only two lan-

guage types with respect to the *pro*-drop parameter: Null-sub-

ject languages (Italian type) and non-NSL (English type).

However, as discussed earlier in this chapter, Rizzi (1982)

predicted the existence of a "mixed" language type. German,

is one such semi-*pro*-drop language. It does not omit thematic

subject pronouns, as shown in (18), though it does permit (re-

quires) expletive subjects to be null, as shown in (19).[15]

---

[15]   In fact, there are probably two types of semi-*pro*-
drop languages, namely those which allow only non-argumental
subjects (excluding subjects of so-called weather verbs which
are quasi-argumental according to Chomsky 1981), for instance
German, and those that allow both non-argumental and quasi-

(18) er/*pro tanzt gern
'he likes to dance'

(19) P geht ins Dorf, weil pro/*es dort getanzt wird
'P goes to the village, because pro there is dancing'

A licensing/identification approach à la Jaeggli/Safir predicts the difference between the German type (semi-pro-drop) and Italian type (full pro-drop). It makes sense that within such a scheme, empty categories which lack semantic content need to meet only the licensing criterion, since there is no content which needs to be recovered via identification. Expletive subjects, by definition, lack any semantic content, and therefore only need to be licensed. In such a system then, the distribution of null-subject types in German can be understood as a result of formally licensing but not identifying null subjects.

The licensing/identification theory might be understood as breaking up the originally monolithic pro-drop parameter into two distinct parameters, which can be paraphrased as a "Null-Referential-Pronoun" (NR) parameter and a "Null-Exple-

---

argumental subjects, for instance Icelandic and Yiddish. Unless otherwise stated, for our purposes both non-argumental and quasi-argumental subjects are subsumed under the term "expletives". Rizzi (1986: 541) lists the following possibilities:
a: no occurrence of pro
b: pro=non-argument (German)
c: pro=non-argument and quasi-argument (Yiddish, Icelandic)
d: pro=non-argument, quasi-argument, and referential argument
(Italian)
Compare the following examples from Icelandic and German:

(i)   hefur pro rignid   (ii) hat es/*pro geregnet
      'has (it) rained'   'has it rained'

tive" (NE) parameter. However, as already implied in Rizzi (1982), these two parameters are **not** completely independent, i.e. not all four theoretically possible combinations of the two binary parameters should actually exist.[16] A combination of [+NR,-NE], that is a language with null referential and overt expletives, is unexpected. Expletives by definition lack any referential function, so the only reason for their occurrence must be that a language does not license null subjects at all. In other words, it is much easier for an expletive to be null than for a referential pronoun to be null, since there is no pronominal content which must be recovered.[17]

There are two main problems with the licensing condition as formulated by Rizzi (1986) and the morphological uniformity hypothesis: First, neither encodes the longstanding intuition that there is some sort of connection between rich agreement and the occurrence of null subjects, and second, it is not clear **why** languages like Italian and Japanese pattern together with respect to null-subject phenomena (if they really do). Although the remainder of this study will focus on the applicability of the licensing/identification approach to Slavic, it is worth summarizing a recent proposal by Speas (1994)

---

[16]   See pages 22-23 for a discussion of Rizzi (1982).

[17]   Within this framework it makes no sense to talk about identification without licensing, since, if null subjects in some language were identified without being licensed, they still could not exist, so that the fact that they were identified would have no consequences.

which attempts to remedy these shortcomings. As theoretical background, Speas assumes familiarity with Pollock's (1989) split INFL hypothesis according to which INFL is decomposed into (at least) two separate functional heads AGR and T, each with its own phrasal projection. This yields the clause structure in (20b).[18]

(20a)   (20b)

```
        CP                                  CP
      /    `                              /    `
   Spec     C'                         Spec     C'
           /                                   /    \
          C      IP         ➡             C      AGRP
               /    `                          /    `
             NP      I'                     NP      AGR'
                    /  `                           /    \
                   I    VP                       AGR      TP
                                                        /    `
                                                     Spec      T'
                                                              /  \
                                                             T    VP
```
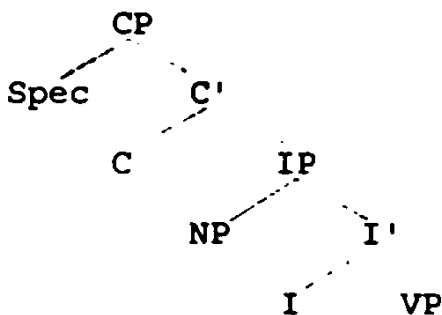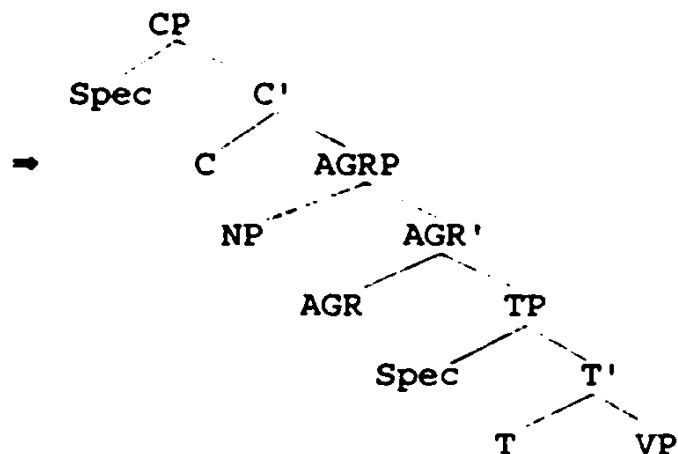
Speas's (1994) proposal aims at **explaining** Jaeggli and Safir's generalization. Her claim is that the null-subject parameter follows directly from the principle of economy, according to which a projection XP can be projected only if it has content. A node has content iff it dominates a distinct phonological matrix or a distinct semantic matrix. Null arguments occur wherever general principles of economy permit them,

---

[18]   I am noncommittal as to the respective position of AGRP and TP, since this has no bearing on the current study. In fact their position might be subject to cross-linguistic variation. The theory of functional projections has received much attention recently, and several other functional projections have been postulated, such as Negation Phrase, Focus Phrase and Determiner Phrase.

and no special licensing conditions on *pro* are necessary. Adapting a proposal by Rohrbacher (1995), Speas proposes that in languages with "strong" agreement, each agreement morpheme has its own lexical entry, while in languages with "weak" agreement, the morphemes do not have independent entries. Verbs in weak agreement languages are listed in the lexicon in verbal paradigms. According to Rohrbacher (1995: 356) "strong" agreement, and therefore lexically listed affixes, are found only in those languages which meet the following conditions.

(21) Conditions for lexically listed affixes:

AgrS is a referential category with lexically listed affixes in exactly those languages where regular subject-verb agreement minimally distinctively marks all referential AgrS-features such that in at least one number of one tense, the person features [1st] and [2nd] are distinctively marked.

The functional head AGR has no semantic content. Therefore, if projected it must have phonetic content. In languages with strong agreement, an agreement morpheme heads its own functional projection AGRP. Such an AGRP has content without its Spec position being filled by an overt NP. In weak AGR languages, on the other hand, the affix is base-generated on the verb, and so [Spec, AGRP] has to be filled in order to give AgrP content. The former situation accounts for null-subject languages, the latter for overt subject languages. The principle of economy of derivation also accounts for the Asian null-subject type. Speas suggests that the AGRP projection is necessary only in languages with some sort of agree-

ment. In languages with no agreement whatsoever, there is no AGRP at any level, therefore the language has null subjects since nothing forces movement into the specifier of AGRP.[19] For Speas, there is no null-subject parameter *per se*; instead, null-subject properties of a language are accounted for by an AGR strength parameter coupled with general principles of economy of representation.

2.4. Null-subject diagnostics

Ideally, we would have only to look at a sentence involving a pronominal subject in order to determine whether or not a particular language belongs to the null-subject type. However, many (if not all) languages can omit various elements, including subjects, in various styles of discourse. English certainly should not be analyzed as a null-subject language. Nevertheless we find English sentences without overt subjects, for instance when recounting a sequence of events in "diary" style or in certain questions.

(22) a. Woke up. Fell out of Bed. Dragged a comb
across my head.

b. Got milk?

Similarly, in German a pronoun can be dropped under certain contextual conditions, provided it occupies the initial

---

[19]   It is not clear whether it makes sense to speak of "null subjects" in this context at all, since no structural subject position [Spec,AGRP] is available.

position in a root clause, or, in other words, the [Spec,CP]

position (cf. Ross 1982 as cited in Huang 1984: 555/56).

> (23) A: Weißt du schon, daß Peter wieder zurück ist?
>   'do you know already that P. is back again'
>
> B: Ja, ja. e Hab ihn schon gesehen.
>   'yes,yes. e have already seen him'

The discourse related omissibility of elements, called ellip-

sis, varies from language to language, and defining the exact

conditions under which such "loss" of lexical material can

happen lies outside the scope of this work.[20] In some lan-

guages, however, including Russian, missing subjects due to

ellipsis can occur in such a wide range of conditions that it

becomes difficult to differentiate between genuine *pro*-drop

and ellipsis.[21]

It follows that one cannot readily determine the status

of missing subjects in a language, and it has been suggested

that we should instead rely on a set of syntactic tests that

can be used as diagnostics of *pro* and therefore the null-sub-

---

[20]   There are several interesting questions which I do
not plan to discuss here: Should we posit an empty category
in the place of the deleted element? If so, which type of
empty category should be posited? For a very good discussion
of the issues involved see Huang (1984). He analyzes the
missing pronoun as a variable in argument position bound by a
zero topic, i.e. referring to the discourse topic. The zero
topic in turn is discourse-bound to the initial topic of a
topic chain. Huang clearly differentiates genuine zero pro-
nouns and variables bound by zero topics. He proposes the
existence of two independent parameters: [±zero-topic] and
[±pro-drop], with four logically possible types of language.

[21]   For a more detailed discussion see chapter 3.

ject status of a particular language (Lindseth and Franks (1994); Franks (1994)).

Typical characteristics of null-subject languages are given in (24).[22]

(24) a. **Only null** pronominals are stylistically unmarked.

b. **Only null** pronouns can function as bound variables.

c. **Only null** 3rd plural pronominal subjects can have arbitrary reference.

(24a) expresses the well-known fact that non-emphatic pronominal subjects are regularly omitted in canonical null-subject languages, while overt pronominal subjects are emphatic. Criterion (24b) has been adopted from Montalbelli (1984), whose analysis of pronominal binding in null-subject languages revealed that empty pronouns do not behave in the same manner as their overt counterparts. In particular, he discovered that there is a curious asymmetry between the interpretation of overt and empty pronouns. Montalbelli (1984) claims that in the following type of construction in a null-subject language an overt pronoun in the same position as *pro* cannot be interpreted as bound, whereas the empty pronoun has to be interpreted as bound:

---

[22] Technically speaking, we have to differentiate between two types of null subject languages:
a) languages which license and identify null subjects
b) language which just license null subjects
In the following I use the term (canonical) 'null-subject language' synonymously with type a), unless otherwise stated.

(25) a. muchos estudiantes piensan que **ellos** son
inteligentes
'many students think that they are smart'

b. muchos estudiantes piensan que *pro* son
inteligentes

I do not completely agree with Montalbelli that the **only** interpretation which is possible for (25b) is a bound reading:

(26) (many x: x a student) x thinks x is intelligent

Instead, a referential (free) interpretation of the empty pronoun is also possible, even though this is not the most likely interpretation, especially if a sentence is given to a native speaker without any context. However, what is most crucial for our discussion, the bound variable reading in (26) is **not** available for (25a). In English, on the other hand, the interpretation of the pronoun *they* is ambiguous between a bound and free interpretation. Montalbelli emphasizes that his "overt pronoun constraint" is only valid for positions where the overt/empty alternation is available. The interpretation of the overt possessive pronoun in sentence (27) is ambiguous in both Spanish and English. In Montalbelli's account this is expected, since the overt possessive pronoun does not have an empty counterpart.

(27) muchos estudiantes piensan que **sus** bicicletas son
azules
'many students think that **their** bicycles are blue'

A version of Montalbelli's overt pronoun constraint can then
be defined as follows:

(28) Overt pronouns cannot be bound if the option
overt/empty is available.

Criterion (24c) is derived from Jaeggli's (1986) analysis
of the so-called arbitrary plural construction. Jaeggli points
out that this type of construction occurs in both English and
Spanish, indicating that its existence is independent of the
setting of the null-subject parameter.

However, while in English it is the overt pronoun *they*
which functions as the arbitrary plural pronoun, as illustrat-
ed in (29), in Spanish **only** a null pronoun can be interpreted
as arbitrary in reference. According to Jaeggli (1986) an
overt pronoun in Spanish can only be interpreted as non-arbi-
trary. In (29) the lexical pronoun *they* is ambiguous between
a specific referential reading (a group of more than one per-
son) and an arbitrary reading (someone; one or more people).
In Spanish, there is again a division of labor between overt
and null pronouns, although perhaps less clearly so than stat-
ed by Jaeggli. Examples (30a) and (30b) below are cited from
Jaeggli (1986). Some Spanish informants were less willing to
interpret (30b) with an arbitrary reading, but all agreed that

(30a) certainly **cannot** have an arbitrary reading. The same results obtained in the constructed examples (31a) and (31b).[23]

(29) they don't allow alcohol on campus

(30) a. ellos llaman a la puerta
'they call$_{3pl}$ at the door'

b. *pro* llaman a la puerta

(31) a. aquí ellos no permiten las bebidas alcohólicas
'here they don't allow alcohol'

b. aquí *pro* no permiten las bebidas alcohólicas

The pronoun *ellos* in (30a) and (31a) must have a specific group of people as reference, while the null pronominal in (30b) and (31b) is ambiguous between definite reference and arbitrary reference. Jaeggli unifies this property of null-subject languages with Montalbelli's analysis, proposing that the overt pronoun constraint and the constraint which does not allow an overt pronoun to have arbitrary interpretation in null-subject languages, are reflecting the same underlying principle. In contexts where the overt/empty alternation is available, overt pronouns must be specified as [-dependent], that is independent in reference, while empty pronouns are [+dependent], fixing their reference by linking to an antecedent

---

[23]   Part of the problem seems to be that Spanish has another tensed construction with arbitrary null subject, the so-called impersonal *se* construction, which native speakers tend to prefer over the arbitrary plural construction dis-cussed here.

through some sort of binding mechanism.[24] As for the arbi-
trary plural construction, a [+dependent] third-person pronom-
inal is assigned the *arb* index at LF if it lacks an anteced-
ent. Jaeggli's constraint on the arbitrary interpretation of
overt pronouns is then defined as follows:

(32) Overt pronouns may not be arbitrary in reference
iff the overt/empty alternation obtains.

---

[24] This generalization raises various problems, the
solution to which lies outside the scope of this work.
Jaeggli is not very specific as to what kind of binding rela-
tion "in an extended sense of the term" is involved for [+de-
pendent] elements to fix their reference. For instance, in
Spanish any null pronominal in subject position is specified
[+dependent]. Thus, in the following sentence the null pro-
nominal has to fix its reference by linking to an antecedent,
supposedly by being bound by AGR located in the finite INFL.

(i) *pro* compró un pulpo
'he/she bought an octopus'

AGR in this instance contains only person and number fea-
tures. Therefore, when uttered outside any context the sen-
tence is ambiguous between *he* and *she*, so that coindexation
with some "outside" antecedent seems necessary to fix refer-
ence to gender. Furthermore, [-dependent] pronouns also seem
to fix their reference via "some kind of binding relation" to
an antecedent.
     Also consider the problem which arises in the following
type of constructions: (example from Montalbelli 1986: 140)

(ii) muchos carniceros creen que *pro* compraron un pulpo
'many butchers think that *pro* bought an octopus'

Here the [+dependent] pronominal subject has to act as a
bound variable, being bound by 'many butchers'. However, it
seems that at the same time it will be bound by the AGR of
its own clause, for the same reasons as in the above example
(i). A third connected question is whether or not empty 1st
and 2nd person subject pronouns in null subject languages
should also have the specification [+dependent] or not. Note
that for them, being bound by INFL does suffice to fix their
reference.

While we are not concerned about the details of Jaeggli's
analysis, it is important because it constitutes the basis
for the third of our null-subject diagnostics. In summary,
the characteristics under (24) can be taken as fairly reli-
able diagnostics as to whether or not a language belongs to
the canonical null-subject languages. In the next chapter
these diagnostics will be applied to Slavic.

.

3. NULL-SUBJECT CHARACTERISTICS OF SLAVIC LANGUAGES

3.1. Application of null-subject diagnostics

The last section of the previous chapter presented a set of three diagnostics (repeated here as (1)) to determine the canonical null-subject status of a language.

> (1) The following properties hold true for null-subject languages:
>
> a. **Only null** pronominal subjects are stylistically unmarked.
>
> b. **Only null** pronouns can function as bound variable.
>
> c. **Only null** 3rd plural pronominal subjects can have arbitrary reference.

The next step is to apply these diagnostics to Slavic. We shall take Russian as representative for East Slavic, Serbo-Croatian for South Slavic, and Czech for West Slavic.

3.1.1. West Slavic and South Slavic

By these criteria most West and South Slavic languages are canonical null-subject languages.[1] Both Czech and Serbo-Croatian do not normally lexicalize pronominal subjects of tensed clauses, as illustrated in (2) and (3). The use of overt pronouns is generally reserved for contrast, emphasis, or a change in discourse topic, shown in (4) and (5).

---

[1]   For Sorbian, which does not behave like a canonical null-subject language, see Chapter 5.

(2) Často *pro* vstávám brzo   [CZ]

(3) Često *pro* ustajem rano   [SC]
     'often (I) get up early'

(4) Alena vstává brzo, a **já** často vstávám pozdě   [CZ]

(5) Jovan ustaje rano, a **ja** často ustajem kasno   [SC]
     'A/J gets up early, but I often get up late'

Diagnostics (1b) and (1c) provide further evidence that Czech and Serbo-Croatian are null-subject languages. The Czech examples under (6) and Serbo-Croatian examples under (7) pertain to (1b), demonstrating that only null pronouns can have bound variable readings.

(6) a. každý student si myslí, že *pro* dostane jedničku

    b. každý student si myslí, že **on** dostane jedničku

(7) a. svaki student misli da će *pro* dobiti desetku

    b. svaki student misli da će **on** dobiti desetku
      'every student thinks that he will get an A'

The bound variable reading rendering interpretation (8) can only obtain in examples (6a) and (7a).[2]

(8) ((every x: x a student) x thinks x will get an A)

---

[2]   The fact that the overt pronoun in the example below may be bound to the quantifier phrase rendering interpretation (ii) is not a contradiction to what has been said so far. In the position after a preposition there is no alternation between empty and overt pronouns and hence Montalbelli's overt pronoun constraint does not apply.

(i)   každí student chce, aby Maria hrála s ním

(ii) ((every x: x a student) x want that M. play with x)

Examples (6b) and (7b), which have overt pronouns, can only be understood in the sense that every student thinks that a particular person is getting an A.

Examples (9) and (10) provide an application of diagnostic (1c). In Czech and Serbo-Croatian, the arbitrary reading obtains only if the subject is 3rd person *pro* instead of overt *oni*.

(9) a. *pro* v tom obchodě prodávají kávu   [CZ]

    b. **oni** v tom obchodě prodávají kávu
       'in this store they sell coffee'

(10) a. *pro* ovdje prodavaju kavu   [SC]

    b. **oni** ovdje prodavaju kavu
       'they sell coffee here'

The overt pronouns in (9b) and (10b) necessarily give rise to the [-dependent] interpretation, so that *oni* in those sentences must refer to a specific group of people.

3.1.2. Russian (East Slavic)

It is easy to find Russian examples with zero subjects in tensed clauses, especially in colloquial Russian or artistic texts imitating colloquial style. The following examples were found in stories by Sokolov, Jerofeev, Romanina (Innis 1985).

Missing subjects in the Russian text are marked with #, other

missing material is given in brackets.[3]

> (11) On zamečal, čto # dumaet o nej postojanno, i
> radovalsja, čto # ničego ne xočet ot neě, i značit
> # dejstvitel'no ljubit (eě).
>
> 'He realized that (he) was constantly thinking
> about her, and rejoiced that (he) didn't want
> anything from her, that means (he) really loved
> her.'

> (12) No daže esli by ja pomnil sejčas eě imja, to # ne
> stal by nazyvat' (ego), kakoe vam delo.
>
> 'But even if I remembered her name right now, then
> (I) would not name it, what business of yours is
> it.'

> (13) Ja ej i stěkla skol'ko raz vstavljal, i peč'
> perekladyval,... a ni razu # ne ugostila (menja).
> Den'gi # daět (mne), a nasčět ětogo nol'.
>
> 'I put windows in several times for her, and moved
> the stove, but (she) didn't invite me for dinner a
> single time. (She) gave me money, but nothing ever
> came of it.'

> (14) Budet vremja - i kryšu # počinju. A v tot den' s
> utra stěkla # vstavljaju.
>
> 'If there is time - (I) will fix the roof as well.
> But on that day (I) I'm putting in windows from
> early morning on.'

---

[3] Non-overt subjects are very frequent when clauses
are coordinated:

> (i) on vošěl v komnatu, # sel za stol i # načal rabotat'
> 'he entered the room, sat at the table and started
> working'

I analyze these cases as VP conjunction, and therefore not as
instances of empty/missing subjects. The second and third
verbs "share" the subject with the first verb. Therefore,
sentence (i) should be represented by the structure (ii) and
not (iii).

> (ii) [$_S$[$_{NP}$on] [$_{VP}$[$_{VP}$ vošěl v komnatu] i [$_{VP}$ sel za stol]]

> (iii) [[$_S$[$_{NP}$ on] [$_{VP}$ vošěl v komantu]] i [$_S$ e [$_{VP}$ sel za
> stol]]]

(15) Prokuror - on pobelel ves', # govorju daže gazetu
     perestal čitat'. # Xodit po učastku - nogami
     oduvančiki # topčet.

'The state attorney turned completely pale, (I)
say, and even stopped reading the newspaper. (He)
walks along the plot, with his feet (he) tramples
dandelions.'

(16) Vse govorjat: Kreml', Kreml'. Oto vsex ja slyšal
     pro nego, a sam # ni razu ne videl (ego). Skol'ko
     raz uže, tysjaču raz, napivšis' ili s poxmeljugi,
     # proxodil po Moskve s severa na jug, s zapada na
     vostok ... - i # ni razu ne videl Kremlja.

'Everybody says: The Kremlin, the Kremlin. From
everybody I hear about it, but (I) myself haven't
seen it a single time. How many times already, a
thousand times, drunk or with a hangover, did (I)
walk through Moscow from North to South, from West
to East, but not a single time did (I) see the
Kremlin.'

Several authors (Müller (1989); Kosta (1990); Benedicto
(1994)) have argued that Russian is a canonical null-subject
language. One reason for this judgement is undoubtedly the
wide range of examples with missing subjects which can be
cited. Below I will argue against this claim on the basis of
the null-subject diagnostic tests. I essentially agree with
Franks (1990), who analyzes missing subjects in examples com-
parable to the ones above as being deleted at PF due to some
kind of discourse-licensed ellipsis. Such loss of lexical
material in Russian is clearly not restricted to subjects, and
thus should be treated separately from the pro-drop phenome-
non. (Colloquial) Russian in general has a strong tendency
toward the deletion of lexical material which is recoverable
from the context, so that the "loss" of subject pronouns is

just a part of this larger phenomenon.[4] Examples (11-13), and (16) cited above illustrate deletion of direct and indirect objects. Ellipses of complementizers as shown in (20) and (21) (examples cited from Nakhmovsky and Leed (1988)) and verbs as shown in (17) and (18) are also frequent:[5]

---

[4] Huang (1984) proposed a "discourse-oriented vs. sentence-oriented" parameter. Nichols (1985) suggested that Russian is a very discourse-oriented language. Discourse-oriented languages tend to follow a rule according to which topic NPs are deleted under identity with a topic in the preceding sentence. Example (13), repeated below, illustrates the complexity of this phenomenon:

(i) Ja ej i stěkla skol'ko raz vstavljal, i peč' perekladyval,.. a ni razu (ona) ne ugostila (menja). Den'gi (ona) daět (mne), a nasčět ětogo nol'.
'I put windows in several times for her, and moved the stove, but (she) didn't invite (me) for dinner a single time. (She) gave (me) money but nothing ever came of it.'

The speaker and a female person are the "known entities" or topics of the discourse. The omission of the first *ona* is particularly interesting, since this is actually a case of contrast, where we would expect overt pronouns in a *pro*-drop language: 'I did this, but she didn't do that...'
It is also worth mentioning that most Russian speakers, while allowing missing objects due to ellipsis, do not allow "syntactically active" *pro*-objects, as in Rizzi (1986):

(ii) il bel tempo invoglia a restare
'the good weather invites *pro* [PRO to stay]'
*xorošaja pogoda soblaznjaet ostavat'sja

[5] Similar elliptical constructions with missing verbs or complementizers were rejected by Serbo-Croatian informants. Example (i) equates to (ii) in Russian as (iii) does to (20).

(i) *Knjigu # vam? (dati)
'Should I (give) you the book'

(ii) knigu # vam?

(iii) *Dima došla # Jovan još spavao. (kada)

This indicates that there is not a direct relationship between the ellipsis phenomenon and *pro*-drop properties of a language.

(17) kuda ty (iděš')
     'where do you (go)'

(18) ty (prixodiš') ko mne
     'do you (come) to me'

(19) Dima prišěl (kogda) ja eščě ležala
     'Dima came (when) I still was lying'

(20) u nas novaja sosedka (kotoruju) Natašej zovut
     'we have a new neighbor (whom) they call Natasha'

Here I will not be concerned about exactly which factors allow recoverable lexical material to be deleted at PF, nor will I attempt to determine if a "bound variable and zero topic" analysis à la Huang (1984), as discussed in the previous chapter, would also apply to Russian.[6] The real problem for this work is that (what are in my view) discourse licensed null subjects may appear to exhibit the null-subject properties in (1), giving the (false) impression that Russian is a canonical null-subject language.[7]

---

[6] At first glance it might be tempting to attribute ellipsis of subjects to "rich" INFL. However, this "explanation" fails to account for the deletion of objects, verbs and conjunctions, and once we claim that zero subjects in Russian are part of a larger phenomenon we would also like to give a unified account of the recoverability conditions of deleted constituents.

[7] One might actually want to argue that missing subjects in Russian are not stylistically unmarked, hence failing criterion (1a). Instead, they may be marked, perhaps as "intentionally colloquial". Jakobson (1939 :150) wrote in that context: " ... en russe c'est justement l'omission excessive du pronom de la première personne que Dostojevski éprouve comme une morgue irritante" ('... in Russian it is precisely the excessive omission of the pronoun of the first person which Dostoyevsky felt to be an annoying arrogance').

Since the diagnostic tests boil down to a division of
labor between overt and null subjects in null-subject languages,
one can couch the same conditions in terms of overt subjects.[8]

(1') a. **Overt** subjects are stylistically marked.

    b. **Overt** subjects cannot function as bound variables.

    c. **Overt** subjects cannot be arbitrary in reference.

The important point is that **overt** subjects in Russian
also exhibit the properties under (1) and **not** those under (1').
This means that Russian does **not** exhibit the division of labor
between null and overt subjects which one would expect for a
true null-subject language. In order to avoid misjudgments
based on the occurrence of null subjects in languages like
Russian, the following investigation will focus on the interpre-
tive possibilities associated with **overt** subjects as summarized
above in (1').

First, in contrast to most West and South Slavic languages
as exemplified by Czech and Serbo-Croatian, there is nothing
emphatic about the overt expression of a pronominal subject in
Russian.

(21) a. **ja** ponimaju vas   [Ru]

    b. **já** vas rozumím   [Cz]

    c. **ja** vas razumem   [SC]
       'I understand you'

---

[8]   See Lindseth and Franks (1994).

In Czech and Serbo-Croatian, the presence of an overt subject clearly indicates opposition, contrast or emphasis, while this is not the case for Russian. In neutral literary style, the use of overt pronouns in Russian is often obligatory, and the dropping of a subject is felt as incomplete and elliptical by native speakers. The normal way of saying 'I love Peter' in Russian is *ja ljublju Petra* with the subject *ja*, while in Czech the normal way is *miluje Petra* without a subject. Russian speakers learning Czech or Serbo-Croatian find the regular lack of pronominal subjects very striking and must make a conscious effort to avoid the overuse of overt pronouns. That should not be the case if Russian were a canonical null-subject language. Sentences (22-24) provide more examples of this contrast between Czech and Russian:

(22) a. # mluvil jste s jeho ženou   [CZ]

     b. **vy** govorili s ego ženoj   [RU]
        'you talked to his wife'

(23) a. # došla pro mléko   [CZ]

     b. **ona** sxodila za molokom   [RU]
        'she went for milk'

(24) a. # bydlím v Americe   [CZ]

     b. **ja** živu v Amerike   [RU]
        'I live in America'

Criterion (1b') similarly divides the Slavic languages into East, on the one hand, and West and South on the other. Russian in (25) allows the bound variable reading despite the

presence of the overt pronominal subject, but this option does

not exist for the otherwise identical Czech (6b) and Serbo-

Croatian (7b) examples repeated below as (26) and (27).

(25) každy student dumaet, čto **on** polučit pjatěrku   [Ru]
     'every student thinks that he will get an A'

(26) každý student si myslí, že **on** dostane jedničku   [CZ]

(27) svaki student misli da će **on** dobiti desetku   [SC]

In Serbo-Croatian and Czech the overt pronoun cannot be bound

by the quantified expression, while in Russian (and English,

for that matter), the overt pronoun can render the bound vari-

able interpretation in (8) repeated here as (28).

(28) ((every x: x a student) x thinks x will get an A)

Roughly the same results obtain if the QP is replaced by

a referential NP, as in (29-31).

(29) Ivan$_i$ dumaet, čto **on**$_{i/j}$ poličit pjatěrku   [Ru]
     'Ivan thinks that he will get an A'

(30) Jan$_i$ si myslí, že **on**$_{*i/j}$ dostane jedničku   [CZ]

(31) Jovan$_i$ misli da će **on**$_{*i/j}$ dobiti desetku   [SC]

Using the overt pronoun instead of *pro* in a null-subject lan-

guage therefore leads to a deictic, i.e. non-dependent inter-

pretation, so that the intentional coreference of an overt

pronoun in a subordinate clause with the matrix subject is

impossible, or at best extremely marginal, in null-subject languages.

According to the third criterion, an overt subject pronoun in a null-subject language cannot be arbitrary in reference. It turns out that this test, although less obvious, further supports the proposed division of Slavic languages.[9] The Russian arbitrary 3rd plural construction is normally described as requiring that the subject be phonologically null. However, some Russian native speakers nevertheless confirmed that the examples in (32) can have an arbitrary interpretation.[10]

> (32) a. v Amerike **oni** govorjat po-anglijski
> 'in America they speak English'
>
> b. vo Francii **oni** edjat ulitok
> 'in France they eat snails'
>
> c. na benzokolonkax **oni** prodajut sigarety
> 'at gas stations they sell cigarettes'

Although the constructions may not be normative with respect to literary Russian, the point is that the presence of an overt pronoun does not actually force a non-arbitrary interpretation. In Serbo-Croatian and Czech, on the other hand,

---

[9] In general, this test seems less reliable than the other two null-subject criteria. Native speaker judgements about the possibility or lack of an arbitrary interpretation of an overt pronoun vary, not only in Russian but also in Sorbian, as illustrated in Chapter 4.

[10] See Lindseth and Franks 1994; Sperling 1990.

the presence of an overt pronoun does not allow an arbitrary interpretation, as discussed above.

I therefore conclude that Russian (East Slavic) is **not** a canonical null-subject language, while Czech (West Slavic) and Serbo-Croatian (South Slavic) are. The observation that Russian is different from other Slavic languages with respect to the expression of subject pronominal is not a new one. Jakobson, for instance, wrote (1939: 150):

> Le russe présente deux variantes stylistiques pour dire "je vais /en voiture/": *ja jedu* /avec pronom personnel/ et *jedu* /sans pronom/. De même, en tchèque: *já jedu* et *jedu*. Cependant, il y a là une grande différence entre les deux langues: ...... en russe, c'est la construction à deux termes qui est le type "normal", tandis que la variante à sujet-zéro est un procédé expressif. En tchèque, au contraire, le zéro d'expressivité se rapporte au sujet-zéro, et la valeur expressive s'attache au type *já jedu*."

Interestingly, even some authors who claim that Russian is a *pro*-drop language comparable to Czech do point out that there is a difference between the languages concerning the properties of zero and overt pronouns, respectively. Koktová (1992: 40,41,60), for instance, investigating the variation among "anaphoric expressions" in Czech and Russian sentences

---

" 'Russian has two stylistic variants to say I'm going / by car /": *ja jedu* /with a personal pronoun/ and *jedu* /without pronoun/. The same, in Czech: *já jedu* and *jedu*. However, there is a big difference between the two languages: ... in Russian it is the two-termed construction which is the "normal" type, while the variant with the zero-subject is an expressive mode. In Czech, on the other hand, zero expressiveness is related to the zero-subject, and the expressive value is attached to the type *ja jedu*.'

of the type *Karel₁ řekl, že 0ᵢ přijde* ('Karek said that (he)

would come') / *Miša₁ skazal, čto {0ᵢ/onᵢ} priedet* ('Miša said

that (he)/he would come') writes:[12]

> My major claim is that in Czech, the basic variation
> is between the zero form and the anaphoric NP as
> unmarked anaphoric expressions (the personal pronoun
> *on* being highly marked, occurring only in emphatic
> (emotional contexts) ... On the other hand, in Rus-
> sian, the basic variation is between the zero form
> or the personal pronoun *on* ... Unlike Czech, the
> relation of the primary (zero) and secondary anaph-
> oric expressions in Russian is rather loose; these
> two anaphoric expressions (the zero form and the
> pronoun *on*) are almost freely interchangeable.

---

[12] "Anaphoric" in this context refers to a pronoun or full NP which derives its reference by virtue of being linked to an NP in a preceding clause. The use of an overt ("secondary") anaphoric expression in a *pro*-drop language means that a certain amount of explicitness is necessary to avoid ambiguity. According to Koktová, the major condition for the usage of an anaphoric NP (instead of the primary anaphoric expression which is zero) in Czech is a nonsubject anteced-ent, especially if there is a danger of ambiguity (sentence (i) is a slightly changed version of Koktová's example).

    (i) Pavel₁ Aleně₁ nic nedal, protože **Alena₁** zlobí
        'P. didn't give A. anything, because A. is bad'

The use of an overt pronoun *ona* would presumably be highly marked and only possible in emphatic (emotionally colored) context. Thus, the Czech and Russian examples below are not equivalent. Instead, (iii) corresponds to example (i) above.

    (ii) Pavel Aleně nic nedal, protože ona zlobí   [CZ]

    (iii) Pavel Alene ničego ne dal, potomu čto ona ploxo
          sebja vedět   [RU]

3.2. Why is Russian different?

3.2.1. The intuitive explanation

The conclusion that Russian (East Slavic) does not qualify as a canonical null-subject language, while Czech (West Slavic) and Serbo-Croatian (South Slavic) do, leads inevitably to the question of why East Slavic is different from the other languages. First of all, the question arises whether or not Russian does not license any null subjects. Within Jaeggli and Safir's framework, this would be unexpected. The Russian verbal paradigm has only derived inflectional forms and thus seems to meet the morphological uniformity hypothesis. Therefore null subjects should be licensed in Russian. Indeed, expletives in Russian may (and generally must) remain phonologically empty.[13] None of the verbs in (33-36) assigns a theta-role to its subject. The first two are weather-verbs, (35) contains a so-called raising verb and (36) a passive.

> (33) *pro* temneet
>   getting-dark
> 'it's getting dark'

> (34) zdes' *pro* duet
>   here   blows
> 'there is a draft here'

---

[13] Some dialects allow overt expletives, as can be seen in the following example taken from Woolhiser (1994):

> ono těplo v izbe
> it warm in cabin

> 'it's warm in the cabin'

Overt expletives in Slavic languages are discussed in Ch. 5.

-54-

(35) mne *pro* kažetsja, čto on ne ponimaet menja
me seems that he not understand me
'it seems to me, that he doesn't understand me'

(36) *pro* bylo napisano ob ètom v gazete
was written on this in paper
'They wrote about this in the newspaper'

The problem which arises is how we can determine what it is which blocks the identification of referential null subjects in Russian. Before trying to formalize this though, it is worth looking at the "old" intuition that there is a correlation between the richness of verbal agreement morphology and availability of null subjects in a language. We should be able to find a relevant difference in the overt expression of agreement features between Russian and the languages which represent West and South Slavic in this discussion.

The present tense paradigm of Russian actually has distinct endings for every person/number combination, as does Serbo-Croatian which is a canonical null-subject language. Czech even has a syncretism between the 3rd person singular and plural within one conjugational pattern.[14] The conjugational paradigms for verbs with infinitival *a-* and *i-* stems given in Tables 1 and 2 below illustrate this.

---

[14]   For a discussion on the allowable numbers and types of syncretisms see section 4.5.1.

Table 1: Conjugational Paradigms: a-stems

RU: obedat'   CZ: obědvat   SC: ručati

	RU	CZ	SC
1sg	obedaju	obědvám	ručam
2sg	obedaeš'	obědváš	ručaš
3sg	obedaet	obědvá	ruča
1pl	obedaem	obědváme	ručamo
2pl	obedaete	obědváte	ručate
3pl	obedajut	obědvají	ručaju

Table 2: Conjugational Paradigms: i-stems

RU: govorit'   CZ: hovořit   SC: govoriti

	RU	CZ	SC
1sg	govorju	hovořím	govorim
2sg	govoriš'	hovříš	govoriš
3sg	govorit	hovoří	govori
1pl	govorim	hovoříme	govorimo
2pl	govorite	hovoříte	govorite
3pl	govorjat	hovoří	govore

Therefore, the present tense paradigm of Russian does not give us the desired answer. At first glance, Russian appears even more qualified to license null subjects than Serbo-Croatian. Within the Jaeggli and Safir system, forms with zero endings, such as the 3rd person singular of the Serbo-Croatian verbs above, pose a potential problem. Recall that licensing in this framework is the result of consistency in the entire paradigm: Null subjects are licensed in languages with either only derived or only non-derived verb forms. If a zero within a paradigm of otherwise consistently derived verb forms is truly an affix itself, then the paradigm can still be uniform. However, if the zero represents no ending at all, then the paradigm is derivationally mixed, and null subjects should not be licensed. The question is whether or not the potentially

problematic 3rd person singular forms *ruča* and *govori* in the Serbo-Croatian paradigms are bare stems. It turns out that we can claim in keeping with Jakobson's (1948) "one stem system" that these forms do actually involve morphologically active, or "real" affixes. First, there is evidence that the Serbo-Croatian form *ruča* involves a consonantal zero ending. Assuming that the basic stem is "ručaj-", we can account for the 3rd person singular present tense form by arguing that the consonantal zero affix of the 3rd person causes truncation of the *jot* due to the unstable C+C environment, just as the *m* of the 1st person singular or *t* of the infinitive causes truncation of the *jot*:

$$(37) \quad \text{ručaj} + \text{ti} \quad \rightarrow \text{ručati} \quad \text{(infinitive)}$$
$$\text{ručaj} + \text{m} \quad \rightarrow \text{ručam} \quad \text{(1st sg)}$$
$$\text{ručaj} + \varnothing_{[+cons]} \rightarrow \text{ruča} \quad \text{(3rd sg)}$$

The *jot* of the basic stem is evident in the imperative form *ručaj*. It is possible that such imperative forms really have no endings at all. However, that would be no problem for the morphological uniformity criterion, since the imperative in general seems to behave independently from the null-subject parameter. Note that even in canonical overt-subject languages

such as English, there is no overt expression of the subject
of an imperative.[15]

Within the same system, the forms *govori and govoriš* in
Serbo-Croatian actually derive in the following way:

(38) govori + i   → govori   (3rd sg)

       govori + iš   → govoriš (2nd sg)

Supporting evidence for this analysis, in opposition to assum-
ing "govori + $\emptyset_{[+cons]}$" and "govori + š", comes from verbs like

---

[15]   Feldstein (1987) comes to a comparable conclusion
in his analysis of Polish *czyta* (read 3 sg), while he posits a
"real" zero desinence for the imperative *czytaj* (imperative,
2sg.). Working with Jakobson's (1948) one-stem system, he
posits "czytaj" as the basic stem. It has been considered a
problem for the one-stem system to explain why a zero-ending
causes the *jot* to truncate in one form (3rd sg) and not the
other (imperative). Feldstein proposes that Polish verbal
desinences in the non-past, including zero endings, can be
further segmented into smaller units representing the conjuga-
tion class, person and number. Based on phonological evi-
dence (i.e. the occurrence of interword sandhi) he assumes
that there is a word boundary between the zero unit and the
person and number endings in the imperative. He further
proposes special truncation rules to deal with desinences
which contain zero elements. If a zero is accompanied by a
non-zero unit within the same word boundary, then the zero
remains without any effect on truncation. If there is only a
single zero within the word-unit, the zero desinence behaves
like a vowel. This is the case in the imperative, and conse-
quently the *jot* is not truncated. If there are multiple
zeroes present within the word unit, they have the truncation
effect of a consonant. This is the case in forms like *czyta*.
The behavior of the different "zeroes" is illustrated below
(# signifies a word boundary).

czytaj + Ø + Ø + Ø# : czyta   (C+C; truncation of "jot")

czytaj + Ø # Ø + Ø# : czytaj   (C+V; no truncation)

It is obvious how this analysis could be extended to Serbo-
Croatian as well.

željeti which obviously belong to the same conjugation class as
govoriti:

(39) Želim Želiš Želi   Želimo Želite Žele

The personal endings of this conjugation type are thus {*-im*,
-iš, *-i; imo, ite,* e}, and the basic stems for the two verbs
above are "govori" and "žele" respectively. Following Jakob-
son's rules for combining basic stems with endings, the final
vowel of the basic stem is truncated due to the unstable V+V
environment when the vocalic personal endings are added. The
infinitive ending {*-ti*}, on the other hand, is consonantal,
and no truncation occurs when it is added to the vocalic basic
stem.

I assume that the Czech form *dělá* (3rd singular) also
involves a zero consonantal affix. All the forms of the para-
digm above can be derived from the basic stem "dělaj", pro-
vided that there is a special rule of vowel lengthening in
Czech in C+C environments in the non-past paradigm. This is
illustrated in (40).

(40) dělaj + m   → dělám (truncation; lengthening)

     dělaj + $\emptyset_{[+cons]}$ → dělá   (truncation; lengthening)

     dělaj + í   → dělají

We have seen that there is nothing in the present tense
paradigm of Russian which would account for the lack of iden-
tification of null subjects. However, there is a crucial

difference in the **past tense** formation between East Slavic on the one hand and West and South on the other, which will be discussed below.

Historically, Old Church Slavonic (OCS) was *pro*-drop, while modern Russian is not. The question is, then, what happened over time which could have caused the loss of *pro*-drop in Russian. The answer lies in the development of the past tense(s). OCS had four different past tenses: The aorist and the imperfect, which were simple tenses, and the perfect and the past perfect, which were compound tenses. Many Slavic languages today, including Russian, have only one past tense, which is derived from the perfect tense of OCS.[16] The perfect tense in Slavic was originally formed analytically with a present tense form of *byti* ('to be'), indicating the grammatical categories of person and number, and the short form of the *l*-participle which showed agreement in gender and number with the subject, as shown in Table 3.[17]

---

[16] The decline of the aorist and imperfect took place relatively early in Russian. It is difficult to give an exact time frame, since this process is reflected differently in different manuscripts. Also, the bookish church style seems to have been much more conservative than the secular (delovoj) style, where these tenses become less frequent at an earlier point in time. Ivanov (1990) gives the 13th/14th centuries as an approximate time for the considerable decline of the imperfect and the aorist in Russian.

[17] The dual number is not crucial for the discussion and is not listed in Table 3.

Table 3: OCS Perfect Tense

Sg.		masc.	fem.	neuter		
1st	jesmь	xodilъ	xodila	xodilo		
2nd	jesi	xodilъ	xodila	xodilo		
3rd	jestъ	xodilъ	xodila	xodili		

Pl.

		masc.	fem.	neuter		
1st	jesmь	xodili	xodily	xodila		
2nd	jeste	xodili	xodily	xodila		
3rd	sǫtъ	xodili	xodily	xodila		

Russian lost the original gender distinctions in the plural and the nasal vowels (stǫtъ → soutъ) relatively early (~12th/13th centuries). Otherwise, the system above can be found in early Old Russian texts.

Sentences (41) and (42) are examples of an intact, perfect tense in Russian Church Slavonic manuscripts:[18]

> (41) a. ne **dalъ jesi** kozъlęte (Ostromir's Gospel 1056-7)
>   not gave aux_{2sg} goat
>   '(you) have not given a goat'
>
>   b. věmь jako otъ boga **jesi prišelъ** (ibid)
>   know that from God aux_{2sg} arrived
>   '(I) know that (you) have come from God'
>
> (42) jako že oci **soutъ ustavili** i **zapovědali**
>   as emph. fathers aux_{3pl} established and commanded
>   tako že stvori samъ (Izbornik Svjatoslava 1076)
>   so emph. do$_{imperative}$ self
>   'as the fathers have established and commanded so do yourself'

---

[18] Examples are cited from Ivanov (1990: 332-337) and Avanesov and Ivanov (1982: 92-106).

However, the manuscripts of the 11th and 12th centuries al-
ready exhibit a frequent use of the perfect **without** auxiliary,
of which (43) is the oldest attested example:

(43) glěbъ knjazъ **měrilъ** more (Tmutorokan Stone; 1068)
'prince Gleb measured the sea' (missing aux$_{3sg}$ jestъ)

(44) ježe otьcь **dajalъ** i rodi **sъdajali** (B-bark #9; 1100s)
'what father bequeathed and relatives bequeathed'
(missing aux$_{3sg}$ jestъ aux$_{3pl}$ sotъ)

The following examples taken from 13th and 14th century
manuscripts show the continuing loss of the auxiliary in the
perfect tense.

(45) a. otьcь moi **oumerlъ** (Laurentian chronicle; 1377)
father my died
'my father died'

b. gdě že ne **xodili** ni dědi naši (ibid)
where emph. not went nor granddads our
'where our grandfathers didn't go'

c. vъ se že lěto V. ratъ **počalъ** (ibid)
in this emph. year V. war started
'in this year V. started the war'

(46) ne **reklъ** mi jesifъ (B-bark #3; 14th cent.)
not said me Joseph
'Joseph didn't tell me'

(47) što selъ **pokupili** bojare (Novgorod chart. 1368-71)
that land bought Boyars
'and that Boyars bought the land'

(48) bratъ **dějalъ** nasilie (Novgorod charter 1264-65)
brother did violence
'the brother committed violence'

Several scholars, including Ivanov (1990) and Černyx (1954), point out that this loss of the auxiliary was particularly frequent in the 3rd person with overt subjects. This is reflected in examples (48-51) above, which all have full NPs as subjects. According to Černyx, the Smolensk Charter from 1229 uses the perfect tense of the 3rd person 23 times, but not a single time with the copula. There is a good reason why the loss of the auxiliary seems to prefer the 3rd person. The examples from the manuscripts show that the copula was dropped when the loss of the pronominal features was compensable by the occurrence of an overt subject. If we assume that Old Russian was still *pro*-drop at that stage, then overt subjects (pronouns) of the 1st and 2nd person were only used in emphatic contexts. For the 3rd person, on the other hand, any full NP could act as the non-emphatic subject and thus the instances of sentences with overt subjects in the 3rd person were much more frequent. Although the auxiliary was preserved longer in sentences with 1st and 2nd person subjects, there are also early examples for the loss of the copula in these persons, as illustrated below.

(49) ja **dalъ** rukoju svojeju (Mstislav's charter 1130)
'I gave with my hand'   (missing aux_{1sg} *jesmъ*)

(50) ty **osklabilъ** sę (Antiouchus Pandects 11th century)
'you smiled'   (missing aux_{2sg} *jesi*)

(51) my vašee bratije ne **obiděli** ni **grabili**
            (Bishop Jakov's letter to Riga, 1300)
'we didn't offend or rob our brothers' (missing aux_{1pl} *jesmъ*)

The original meaning of the perfect tense was that of an event which started in the past and which continues into the present. Formally this "bridging of time" was conveyed by the present tense of the auxiliary. According to Avanesov and Ivanov (1982: 92-107), the documents of the 12-14th centuries reveal that this original meaning of the perfect tense was in transition at that time. In this context they cite several examples of formally intact perfect tenses with simple past tense meaning. The gradual loss of the auxiliary as described above may have been both a contributing factor to and a logical consequence of this process. The other purpose of the auxiliary, namely that of expressing the category of person, is perhaps more important for our investigation. As mentioned earlier, this role of the auxiliary was compensated for by the use of an overt subject, NP or pronoun.[19] As the perfect tense became established without the auxiliary, overt pronouns became disassociated from their emphatic function and started to spread widely in Old Russian texts.[20] Ivanov points out

---

[19] Old Church Slavonic did not have a 3rd person personal pronoun. 3rd-person reference was made by the demonstratives *tъ*, *sъ* and *onъ*. The latter developed into the 3rd person pronoun.

[20] The shift in the meaning of the perfect tense cannot be interpreted as the sole reason why East Slavic developed non-emphatic personal pronouns and thus lost its *pro*-drop properties, since the same change happened in other Slavic languages which remained null-subject. It seems that all Slavic languages which use the *l*-participle as the simple past tense form of the verb could have gone one way (keep the auxiliary) or the other (use non-emphatic subjects pronouns), once the "bridging function" of the auxiliary was lost. In Polish, the auxiliary forms of the OCS perfect tense have

that the formally intact perfect is still occasionally found in later manuscripts (14th and 15th centuries), especially in the religious style. Interestingly, we can find such examples of the analytical perfect tense both with and without overt subject, which shows that the use of (non-emphatic) overt pronouns (probably due to the loss of the auxiliary in normal speech) had already gained in popularity.

(52) sego ty že esi xotělъ (Laurentian Chronicle 1377)
     this you emph. aux₂ₛₒ wanted
     'you wanted this'

In summation, it seems plausible to attribute the loss of (referential) *pro*-drop to the reinterpretation of the *l*-participle as a verbal form, which due to its participial origin, is somewhat peculiar in shape: Contrary to other forms of the verbal paradigm, it does not show person, but it does show gender instead.[21]

---

been incorporated into the verb form, functioning as desinences which represent the category of person/number, very much like desinences on present tense verb forms.

> *pytać* ('to ask')
>
> | | present | past |
> |-----|---------|--------|
> | 1sg | pytam | pytałem |
> | 2sg | pytasz | pytałeś |

[21] The fate of the past perfect, the third "extinct" tense, was indirectly influenced by the development of the other past tenses. Originally, the past perfect was formed by the imperfect of the verb *byti* and the *l*-participle (e.g. *nesli běste* 'you-all had carried'). In the 13th century the past perfect started using the perfect tense of *byti* instead of the imperfect (e.g. *nesli jeste byli* 'you-all had carried'), as shown in (i). This is presumably connected to the loss of the imperfect tense at around the same time.

In contrast to East Slavic, West and South Slavic did not drop the auxiliary verb in the formation of the past tense, so that the auxiliary has preserved its purpose of expressing the category of person/number and the personal pronouns their emphatic function, as was true for OCS. This is shown below for Czech and Serbo-Croatian.

(53) Četla **jsem** tu knihu [CZ]
read$_{fem}$ aux$_{1sg}$ this book
'I read this book'

(54) Čitala **sam** tu knjigu [SC]

3.2.2. How to formalize the difference between Russian and West and South Slavic

It would seem plausible that the special development of the past tense would have necessitated the use of overt pronouns in that tense only in Russian. However, thematic null subjects

---

(i) a čto selъ i svobodъ ...
'and as far as their land and freedom ...'

to **dali** **jesme** byli andrěju
that gave$_{1-pl}$ aux$_{1pl}$ aux$_{1-pl}$ Andrej
'we had given it to Andrej' (Novg. charter 1304/05)

At about the same time, the perfect starts to lose its auxiliary, which is also reflected in the past perfect. This is illustrated in (ii).

(ii) a ot nemecъ **byli** na sude **iskali** kolokola
and from Germans aux$_{1-pl}$ at court searched$_{1-pl}$ bell
(1284; Gramota knjazja Fedora Rostislaviča)
'but they had demanded the bell from the Germans in court'

The past perfect in this latter form is preserved until the 17th century, but is missing from contemporary literary Russian.

in general stopped being identified, even in the present tense where unambiguous identification never seems to have been at risk.[22] This may imply that the whole paradigm, and not parts of the paradigm or even individual verb forms, determines whether empty subjects are possible or not.[23] The 3rd person deserves further comment. In Slovene, Serbo-Croatian, and both Sorbian languages the copula is used in all forms of the compound past (perfect) while the other South and West Slavic languages have a zero copula in the 3rd person. This is illustrated below.

(55) ljudje so videli [SLN]
people aux$_{3pl}$ saw$_{pl}$

Janez je videl
Janez aux$_{3sg}$ saw$_{masc}$

(56) ljudi su videli [SC]
people aux$_{3pl}$ saw$_{pl}$

Jovan je video
Jovan aux$_{3sg}$ saw$_{masc}$

---

[22] Recall that following Jaeggli and Safir (1989) we differentiate between "licensing" and "identification" of null subjects. All null subjects have to be licensed, but only thematic subjects have to be identified.

[23] A hypothesis stating that this would be universally true is inconsistent with the *pro*-drop properties of Hebrew (among others). In Hebrew, null referential subjects occur only in the past and future tenses, presumably because the present tense AGR is not sufficiently rich to identify *pro*. The situation is further complicated by the fact that 3rd person referential *pro* is licensed only in embedded contexts with an antecedent for the *pro*-element in the matrix clause. *Pro* in matrix clauses is possible only in the 1st and 2nd persons. For a possible analysis see Borer (1989).

(57) luźe   su   wiźeli   [LS]
     people aux$_{3pl}$ saw$_{pl}$
     Jan jo   wiźel
     Jan aux$_{3sg}$ saw$_{masc}$

(58) lidé   viděli   [CZ]
     people saw$_{pl}$
     Jiří viděl
     Jiří saw$_{masc}$

(59) ludzie widzieli   [PO]
     people saw$_{pl}$
     Jerzy widział
     Jerzy saw$_{masc}$

The verbal agreement feature of person is nevertheless also available in all South and West Slavic languages in the 3rd person: A zero copula signals 3rd person. This shows that richness is not necessarily identical to overt expression of agreement features on a given verb form. In other words, it is not the lexicalization of AGR on an individual verb form which makes an empty subject possible.

It turns out that to formalize what exactly blocks identification in Russian is a fairly difficult task. Since Russian actually licenses expletive subjects, an analysis à la Speas (1994) is bound to fail. Recall the split INFL hypothesis, according to which IP is decomposed into two functional projections: AGRP and TP. Speas further argues that, according to the principle of economy, a projection XP can be projected only if it has content. Since AGR has no semantic content, AGRP can be projected under two conditions: The head itself is

filled by an agreement affix, or the specifier position is filled.[24] Returning to the Russian problem, we could assume, adjusting the definition of "strong" agreement appropriately, that Russian is a language with "weak" agreement, due to the lack of agreement morphology in the past tense.[25] The verbal affixes would then not have independent lexical entries, but instead be listed in the lexicon in the verbal paradigms. In keeping with the principle of economy, the specifier of AGRP would thus have to be filled, disallowing *pro*. However, given this hypothesis, we could not account for the occurrence of null expletives, unless we were to assume that this phenomenon

---

[24] The 3rd person forms without copula in West Slavic might be problematic for Speas's analysis. According to Speas (1994: 8) the relevant notion of content is the following:

A node X has content if and only if X dominates a distinct **phonological** matrix or a distinct semantic matrix.

Forms like Czech *četl* ('he read') or Polish *czytał* do not have an overt person/number agreement affix (I assume that -*ł* is the head of TP). Consequently, following the economy of projection principle, we would be forced to assume that AGRP is not projected at all when these forms are being used, since the AGR does not dominate a phonological matrix and there is no subject occupying the specifier position. However, this theory contradicts the intuition that the agreement features are really present and available, though not **phonologically** present.

[25] Under Speas's criteria (using Rohrbacher's (1995) "conditions for lexically listed affixes", cf. Chapter 2) Russian could also be analyzed as a "strong" AGR language, thus theoretically allowing null subjects. However, Speas also mentions explicitly that a "strong" AGR language may choose to list agreement affixes as part of the verbal paradigms, thus forcing pronominal subjects to be overt. While this also raises the question of how to account for null expletives, the **main** problem with Speas's analysis, from my perspective, is that it is not specific enough to say anything interesting about **why** Russian is different (See Chapter 4, fn. 35 for a discussion of Sorbian within Speas's system).

in general were due to some factors unrelated to the issue of *pro*-drop.

In Franks (1990), Lindseth and Franks (1994), and Franks (1994) an analysis within the licensing and identification approach is proposed. Russian fulfills the morphological uniformity criterion and therefore licenses null subjects. The lack of identification in Russian can be formulated in terms of an anomaly in the functional projections of V. The analysis corresponds to the intuition that the acceptability of *pro* should indeed be formulated in terms of richness of inflection. We have seen that although Slavic verbs may exhibit both person-number agreement (AGR_{pen}) and gender-number agreement (AGR_{gnd}), this is realized differently in East Slavic, compared to West and South Slavic. Whereas AGR_{pen} is consistently marked in South and West Slavic, in East Slavic it is absent in certain forms. Specifically, there is no present tense copula, and hence, as we have seen above, all verbs in the past tense lack AGR_{pen} agreement morphology. In East Slavic, the two agreement types are mutually exclusive, which, in principle, would allow us to associate either type with what is generally known as the AGR (or AGR_S) projection in the literature. However, AGR_{pen} and AGR_{gen} in West and South Slavic can be present at the same time (such as in the past tense) and occupy distinct positions. The former is a relatively

free clitic, while the latter appears unequivocally as an ending after the tense suffix -*l*.[26]

This constitutes evidence that in a full-blown functional projection system the two agreement types should be two distinct heads. The Slavic clause structure could then be illustrated by the following tree structure in (60), whereby the relative order of the different functional projections may vary from language to language.

(60) CP

C AGR$_{pen}$P

Spec AGR$_{pen}$'

AGR$_{pen}$ AGR$_{gend}$P

Spec AGR$_{gend}$'

AGR$_{gend}$ TP

Spec T'

T VP
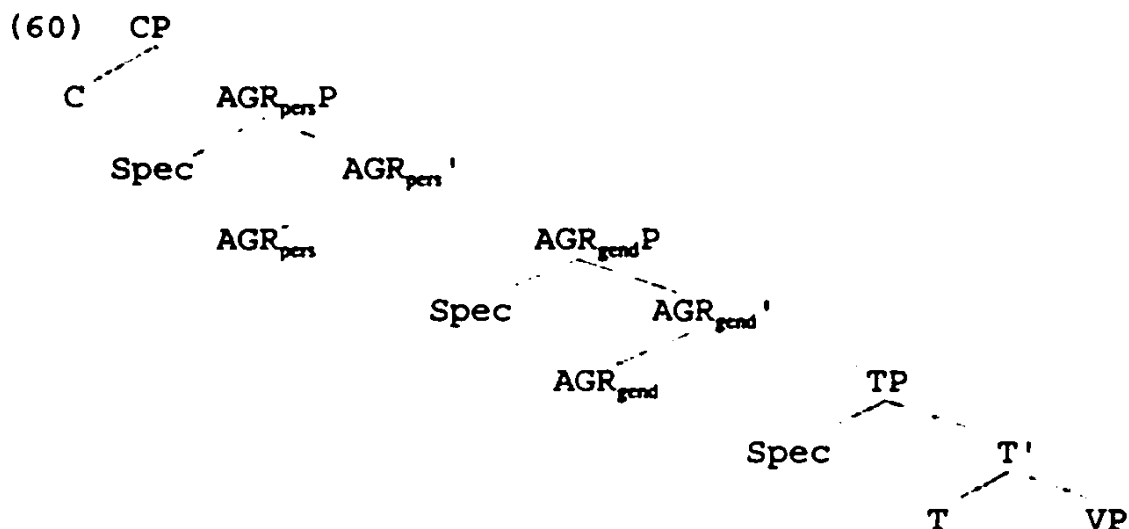
Only AGR$_{pen}$P seems to be relevant for the identification of referential null subjects. In West and South Slavic languages null subjects are identified despite the lack of AGR$_{gend}$ in all forms of the present tense. Thus, a correlation between the availability of a null referential *pro* subject and the presence of AGR$_{pen}$ is proposed. Under this analysis AGR$_{pen}$P is

---

[26]   Even in Polish, the personal endings of the Past tense can be transferred and attached to other words in the sentence:

Gdzie byłaś?   or   Gdieś była? ('where were you')

associated with what is usually called AGR$_S$P. The Russian present tense forms show full person agreement, and yet *pro* is no more appropriate in present tense sentences than it is in the past tense, which lacks AGR$_{pers}$. It seems to be the absence of AGR$_{pers}$ in the past tense which is responsible for blocking the identification of referential null subjects in general in Russian. It should be stressed once more, as already indicated above, that what is important under this analysis is the role of AGR$_{pers}$ in the inflectional *system*, rather than its presence in individual forms.

3.3. Summary

In this chapter I applied three null-subject diagnostics to the Slavic languages. By these criteria, Czech and Serbo-Croatian, as representatives of West and South Slavic, are canonical null-subject languages, while Russian is not. The reason why Slavic languages divide along this line is found in the different development of the past tense. Specifically, East Slavic lost the present tense copula, which led to a loss of person agreement in the past tense. The loss of the auxiliary started in sentences with overt subjects (full NPs or emphatic 1st and 2nd person pronouns), but eventually spread much further and was compensated for by the use of overt pronouns in general, which thereby lost their emphatic character. As a result, the division of labor between overt and zero pronouns, which is characteristic for null-subject languages,

was lost. This insight was formalized by proposing two different functional projections $AGR_{pers}P$ and $AGR_{gend}P$ for the different types of agreement in Slavic. Only the former is true subject agreement in the traditional sense (AGR_S) and relevant for identification of null subjects. The **inconsistent** realization of AGR_{pers} was found to be the culprit for the blocking of identification in Russian.

4. IS SORBIAN A *Pro*-DROP LANGUAGE?

The null-subject diagnostics previously applied to other Slavic languages produce the unexpected result that Sorbian is not a canonical null-subject language. Three possible explanations for this are explored in section 4.6. While all three are workable in principle, the last of these, based on a competition model of language change (Sprouse & Vance 1995), proved to present the fewest problems and give the most plausible account for the whole range of Sorbian data. This is partially due to the fact that it incorporates the broad historical context of the language.

4.1. The Sorbian population of Germany

Soon after the reunification of Germany, in 1991, a political scientist from Bonn wrote:

> Sorben sprechen fließend deutsch, sie besitzen deutsches Wahlrecht und deutsche Pässe, sie sind deutsche Staatsbürger -- aber sie sind keine Deutschen! Deutsche Kaiser, Führer und General- sekretäre haben versucht, ihnen die Luft abzu- drücken -- mal im Würgegriff, mal in der Umarmung. (Oschlies, 1991: 2)[1]

In general terms this 1991 commentary contains some basic points which characterize any study of the Sorbian population

---

[1] 'Sorbs speak German fluently, have German voting rights, German passports, are German citizens -- but they are not Germans. German emperors, "Führer" and General Secretaries have tried to suffocate them -- sometimes in a stranglehold, sometimes in an embrace.'

of Germany, and as will be evident below, this is also relevant for the field of linguistics.

The Sorbs today are a national minority in a very small piece of territory called die Lausitz (Lusatia) in what was formerly the German Democratic Republic (East Germany). In English they are known variously as Sorbs, Lusatians or Wends. In this work I will follow the current style and refer to the language as Sorbian, the people as Sorbs, and the territory as Lusatia.

The Sorbs refer to themselves as *Serbja* (*Serby* in Lower Sorbian) and are the last survivors of the Slavonic tribes which once occupied most of the territory between the Oder and Elbe rivers. They are Slavs, but unlike other Slavic nations they have no separate political existence and are not a subgroup of any larger ethnic identity outside Germany. They want to be thought of as a Slavic people -- as distinct as the Russians, Bulgarians, or Poles -- "minor" only in a quantitative rather than qualitative sense. However, even in Germany there are many people who have never heard about the Sorbs. In fact, Western tourists in eastern Germany were known to panic when they entered the bilingual Lusatian region, believing that the Sorbian language place-name signs meant they had unwittingly crossed the border into Poland. Moreover, even scholars who have specialized in the study of the Czech and Polish languages admit to knowing little about the Sorbs, some having been taught that Sorbian was a dead language.

The area where one can still hear Sorbian spoken current-
ly begins about 50 miles south-east of Berlin and stretches
approximately 55 miles to the south. It is about 40 miles
across at its widest point with Cottbus being the center of
Lower Lusatia (so named because it is downstream on the River
Spree) while Bautzen is the principal focus of Upper Lusatia.

The modern regions of Upper and Lower Lusatia correspond
roughly with the areas occupied by the Milzane and Lusizi
peoples, respectively, prior to German immigration during the
10th century, and the historical delineation of two Sorbian
literary languages appears even in the earliest translations
of religious works during the 16th and 17th centuries. However,
it was German political development in the 17th century which
finally separated these regions and encouraged the emergence
of two major cultural and economic centers: Cottbus in Branden-
burg/Prussia and Bautzen in Saxony. As a result, the dialects
spoken in these two places were codified as the modern liter-
ary languages.

It is difficult if not nearly impossible, to find reli-
able information about the number of Sorbian native speakers.
The Communist government of East Germany was supposedly not
interested in research which could establish such data, since
the results might denigrate the officially proclaimed
"sozialistische Nationalitätenpolitik" by revealing the pro-
gressive assimilation of the Sorbs. According to a 1987 pro-
jection by the *Institut für sorbische Volksforschung*, the

number of people who knew Sorbian was estimated at about 67,000 (16,200 of which were speakers of Lower Sorbian). However, this figure included a large number of persons with only very limited language skills, individuals who learned Sorbian in school as a foreign language, and some who did not use it at all as a language for everyday communication. Recent study indicates that it is only in the comparatively tiny Catholic area of rural Upper Lusatia (to the north-west of Bautzen and roughly defined by the villages of Radibor, Crostwitz, Panschwitz-Kuckau, Ralbitz and Räckelwitz) that one can still find predominantly Sorbian-speaking communities. According to Elle (1991), the 15,000 Sorbian speakers of this Sorbian "Kerngebiet" (core area) make up over ¾ of the total number of people who actively use Sorbian today and thus could be considered native speakers. It can then be deduced that the total number of native speakers of Sorbian must be around 22,500. According to Spieß (1995) there are about 5,100 native speakers of Lower Sorbian, which leaves about 2,400 native speakers of Upper Sorbian outside the Catholic "Kerngebiet". Elle (1992: 73-79) points out that in the "Kerngebiet" 96% of the people under 35 who know Sorbian learned it at home, 95% speak Sorbian in their own families, and 76% speak only Sorbian (never German) at home. In the other (much larger) region of Lusatia only about 29% of the people under 35 who know Sorbian learned it at home, and only 9% use it today in their families. This is somewhat surprising, considering that

over 90% of the ethnic Sorbs in that region have at least some degree of knowledge of Sorbian. Also, there is a significant difference between generations, which is not the case for the "Kerngebiet": 49% of the people over 59 who know Sorbian still use the language at home, while for the group between 36 and 59 this figure falls to about 22%.

The numbers given above show clearly that most regions of Lusatia have succumbed to German assimilatory pressure and are German dominated mixed-language territories at best. This mixture is hardly surprising since German cultural pressure has been a constant force since the 10th century, though at that time Lusatia still had substantial contact with her Polabian neighbors to the north, the Bohemians to the South and Poles to the East. Continued German expansion, however, meant that by the 15th century the Sorbian speech area was surrounded almost entirely by German speakers. Despite almost 1000 years of Germanization, however, Sorbian is by no means "dead". The Sorbian publisher Ludowe nakładnistwo Domowina still publishes a daily Upper Sorbian and weekly Lower Sorbian newspaper, as well as the academic journal *Lětopis* and the pedagogical journal *Serbska Šula*. There is also a significant number of contemporary popular and children's books published every year, and the academic and reference works by researchers of the Sorbische Institut in Bautzen and the Institut für Sorabistik at Leipzig University are among the most complete for any Slavic language. There are also Sorbian kindergartens and

schools, and a Sorbian-language radio station affiliated with Mitteldeutsche Rundfunk.

4.2. Verbal Morphology of Upper Sorbian

(Upper) Sorbian is among the more archaic of the modern Slavic languages. For instance, it has preserved the use of the Dual number, and the Aorist and Imperfect tenses from Common Slavic. The present, perfective future, aorist, and imperfect are expressed synthetically. The aorist is formed from perfective verbs and the imperfect from imperfective verbs. The endings are the same for both tenses with the exception of the 2nd and 3rd person singular, and several grammars treat both tenses together as the simple past tense. The perfect, past perfect, and imperfective future are expressed analytically. The perfect tense is formed with the present tense of the verb *być* 'to be' and the *l*-participle, which is analogous to Old Church Slavonic (OCS). Similarly, the past perfect tense takes the imperfect of *być* and the *l*-participle.

Table 1: Verbal Paradigm of Upper Sorbian

dźěłać = 'to work'

	Present/Future	Imperfect	Aorist
Sg.			
1.	(na)dźěłam	dźěłach	nadźěłach
2.	(na)dźěłaš	dźěłaše	nadźěła
3.	(na)dźěła	dźěłaše	nadźěła
Pl.			
1.	(na)dźěłamy	dźěłachmy	nadźěłachmy
2.	(na)dźěłaće	dźěłašće	nadźěłašće
3.	(na)dźěłaja	dźěłachu	nadźěłachu
Dual			
1.	(na)dźěłamoj	dźěłachmoj	nadźěłachmoj
2.	(na)dźěłataj/-atej[2]	dźěłaštaj/-aštej	nadźěłaštaj/-aštej
3.	(na)dźěłataj/-atej	dźěłaštaj/-aštej	nadźěłaštaj/-aštej

l-participle:

	masc.	fem.	neuter
Sg.	(na)dźěłał	(na)dźěłała	(na)dźěłało
Pl.	(na)dźěłali		
Dual	(na)dźěłałoj		

być = 'to be'

		Present	Imperfect
Sg.	1.	sym	běch
	2.	sy	bě(še)
	3.	je	bě(še)
Pl.	1.	smy	běchmy
	2.	sće	běšće
	3.	su	běchu
Dual	1.	smój	běchmoj
	2.	staj/stej[3]	běštaj/-štej
	3.	staj/stej	běštaj/-štej

---

[2]   Virile vs. non-virile distinction.

[3]   Virile vs. non-virile distinction.

Examples of Perfect:

ja sym dźěłała   'I have been working'
mój smój dźěłałoj   'we two have been working'

Examples of Past Perfect:

ja běch dźěłała   'I had been working'
mój běchmoj dźěłałoj   'we two had been working'

Building on the discussion in Chapters 2 and 3 we can see that the Upper Sorbian verbal paradigm meets the morphological uniformity condition, since all forms are consistently derived. Thus, Sorbian should license null subjects. It would also seem that the "richness" of the Sorbian verbal agreement system can easily compare with that of the other West (and South) Slavic languages. In Chapter 3 we established that Czech (as a general representative of West Slavic) qualifies as a canonical null-subject language. Consequently, we would expect the same to be true for Upper Sorbian. Also, since Common Slavic was a *pro*-drop language, and the Sorbian verbal system seems to preserve the Common Slavic situation as well as any of the other Slavic languages, the expectation that Sorbian is a null-subject language is reinforced.

4.3. Grammar vs. Reality

Upper Sorbian has personal pronouns for the singular, plural, and dual.

Table 2: Personal Pronouns of Upper Sorbian

	1. pers.	2. pers.
Sg.	ja	ty
Pl.	my	wy
Dual	mój	wój

	3. pers.		
	masc.	fem.	neuter
Sg.	wón	wona	wono/wone
Pl.	woni, wone		
Dual	wonaj, wonej		

The dialectal distribution between the two neuter forms *wono* and *wone* will be discussed in Chapter 5. The plural form *woni* is used to refer to males, whereby one individual in the group is sufficient to use this form of the 3rd person plural pronoun. The form *wone* is used elsewhere, and can refer to people, animals and inanimate things. The same distinction is expressed in the dual.

According to current school grammars and older influential grammars such as Jordan (1841), Schneider (1853), Liebsch (1884) and Kral (1895), personal pronouns in Upper Sorbian should only be used in emphatic contexts. Šewc (1984: 117), for instance, writes:

Na rozdžěl wot němskeje a ruskeje rěče wužiwa so w
serbščinje forma personalneho pronomena při werbach
jako subjekt jenož we wobmjezowanym rozsahu.
Zwisuje to z tym, zo charakterizuja same wosobowe
kóncowki hižo jasnje dosć wotpowědnu wosobu.

'In contrast to German and Russian, the personal
pronouns in Sorbian are more restricted in their
use. This is connected to the fact that the personal
endings themselves already indicate the correspond-
ing person quite clearly.'

This is certainly reminiscent of the description of personal

pronoun use in null-subject languages such as Czech or Serbo-

Croatian discussed previously.

In most of the older grammars the authors are even more

outspoken about the lack of overt pronoun use in Upper Sorbian.

For example, Jordan (1841: 169) writes:

Da die Personen an dem Ausgange des Verbums hingäng-
lich erkennbar sind, so werden die persönlichen Für-
wörter *ja, ty, wón, mój, wój, wonaj, my, wy, woni*
nur dann gesetzt, wenn die Person besonders ausge-
zeichnet werden soll.[4]

Similarly, Schneider (1853: 250) writes:

... die persönlichen Fürwörter: 'ich, du, er, wir,
ihr, sie' werden im Wendischen nur dann gesetzt,
wenn ein besonderer Nachdruck darauf ruht, oder
wenn ein Gegensatz angezeigt werden soll. Z.B. Ja
sem to pisaw, nic té ich habe das geschrieben, nicht
du.[5]

---

[4] 'Since the person is sufficiently recognizable from
the verb ending, the personal pronouns *ja, ty, wón, mój, wój,
wonaj, my, wy, woni* are only used if the person is to be
especially marked.'

[5] '... the personal pronouns: 'I, you, he, we, you,
they' are used in Sorbian only if they carry a special empha-
sis or if they indicate a contrast. For example, *Ja sem to*

Liebsch (1884: 27/28) gives the same rule even more explicitly:

> Dieses Verhältnis ... wird im Wendischen ähnlich wie
> im Lateinischen und Griechischen durch die
> Personalendungen des Verbum bezeichnet. Deshalb ist
> die nochmalige Bezeichnung desselben durch die
> Personalpronomina durchaus **überflüssig** und nur dann
> notwendig und gestattet, wenn wir das Subjekt aus
> irgend einem besonderen Grund **nachdrucksvoller**
> hervorheben wollen ... Das ausdrückliche Bezeichnen
> der Person durch das persönliche Fürwort im
> Wendischen ist daher nur ein **Germanismus**.[6]

Finally, Kral (1895: 58) writes:

> Die persönlichen Fürwörter als Subjekte werden nur
> dann gesetzt, wenn auf denselben ein gewisser
> Nachdruck liegt, oder wenn sie im Gegensatze
> stehen.[7]

Thus our expectation that Upper Sorbian is a canonical null-subject language seems to be met. This is confirmed by the following example from contemporary literary Upper Sorbian. The excerpt is taken from Jurij Koch's novel *Ha lećała je módra wróna* published in 1991. (# indicates all places where a subject is "missing")

---

pisaw, nic ty ('I wrote that and not you').'

[6]   'This relationship ... is expressed in Sorbian, very much as in Latin and Greek, by the personal endings of the verb. Therefore, the additional expression by personal pronouns is completely **superfluous**, and only necessary and permitted if the subject is supposed to be stressed **more emphatically**. ... The explicit marking of the person by a personal pronoun in Sorbian is thus just a **Germanism**.'

[7]   'Personal pronouns as subjects are only used if they carry a certain emphasis, or if the are used to contrast.'

Moja maćerna rěč je serbšćina ... Z tutej rěču sym
wotrostI. # njemóźu so wjace wot njeje dźělić, teź
nic, hdyź bych # to chcyI. # Njejsym sej ju wupytaI.
Bě tu hiźo, hdyź # dóńdźech.

'My mother tongue is Sorbian ... With this language
I grew up. I cannot part from it any more, not even,
if I wanted to. I did not select it. It was here,
when I arrived.'

However, Upper Sorbian dialects provide us with contra-
dictory data. Published transcripts of dialect texts reveal
a surprising level of overt subject pronoun use, despite the
fact that the dialects have the same rich person/number system
as the standard language. The following are some typical
examples. Sentences (1-3) are taken from Faßke and Michalk
(1989), while (4) and (5) are from Michalk and Protze (1974).[8]

(1)   a   do   pównoce smó   mó khodźili
      and till midnight aux_{1pl} we went[9]
      'we walked until midnight'

(2)   tuijs sej woni źane(j) rade   wejdźili nejsu
      also refl they none   advice knew $neg\text{-}aux_{3pl}$
      'and since they didn't know what else to do'

(3)   ale ja sn   něk tam   wele moli ribach   bów
      but I aux_{1sg} then there many times mushrooms was

      a   ja (ja)c jich ničo   widźiw nejsym
      and I more they$_{gen}$ nothing saw $not\text{-}aux_{1sg}$
      'but I often used to go there for mushrooms and I
      didn't see any of them anymore'

---

[8]   Some of these examples contain dialect forms which
are not found in standard Sorbian.

[9]   The notation "aux_{1sg}" etc. assumes present tense of
the auxiliary *byč* unless otherwise noted.

(4) **woni** praja, **woni** dyrbja tom knjez Šminderej, zo
     they say   they must that$_{dm}$ Mr   Šminder$_{dm}$ that

     woni, hejzo jich njepušća, **woni** dom njepóńdźeja,
     they if   them not-admit they home not-will-go

     dójž joh **woni** njeńdźa   widźeć
     until him they not-will-go see

     'they say **they** must see Mr. Schminder, that **they**,
     if **they** are not admitted, **they** won't go home,
     until **they** have seen him'

(5) hale kóždy mól pak bóch **ja** cedlku přitykwa;
     but every time but aux$_{1sg sub}$ I note   included;

     ja sym jare rjenje pisawa
     I aux$_{1sg}$ very nicely wrote

     'but every time I included a small note; I wrote
     very nicely'

An interesting descriptive discussion of the problem is
provided by Faska (1964). He concludes from an analysis of a
large number of recordings of speakers from various regions
that the traditional "null-pronoun rule" conflicts with actual
use in Sorbian. In fact, in the sample he analyzed, there
were more instances of overt subject pronouns in tensed clauses
than instances of zero pronouns. It is not surprising that
overt pronouns are used in emphatic and contrastive contexts
in spoken Sorbian. However, Faska (1964) also determines
other environments where spoken Sorbian more-or-less consis-
tently uses an overt (non-emphatic) pronoun, and it is these
instances, illustrated in the examples (6-8), which are of
particular interest to us.[10]

---

[10] All examples cited here are from Faska (1964); some
contain dialect forms which are not found in standard Sorbian.

(6) **If direct or indirect speech is being introduced:**

   a. "ja so ničo njeboju", je **wón** rjekł

      'I don't fear anything', **he** said

   b. a ja njeběch tak hłupy, **ja** sym přeco prajił, zo
      ja dale na dźěło póńdu do města

      'but I was not so stupid, I usually said,
      that I would keep on going to town to work'

   c. **woni** su mi prajili, štó je to byl, ale ja sym to
      zabył

      '**they** asked me what that was but I forgot'

(7) **If the matrix subject is a demonstrative pronoun:**

   a. **tón** te dwaj abo tři dny spać njemóžeše,
      dokelž **wón** dyrbješe wšo widźeć a wjesć

      '**that one** couldn't sleep for two or three days,
      because **he** had to see and lead everything'

   b. a **ta** je so tak wustróžała, zo je **wona** chroma
      wostała

      'and **that one** got so startled that **she** stayed
      lame'

(8) **If the subject of the matrix clause is a full NP:**

   a **mój nan** je prajił, zo bdźe **wón** přećiwo Vogtej hić
   'and my father said that **he** would go against the
   overseer'

In most other environments (except for imperatives,
impersonal sentences, and sentences with 3rd person arbitrary
reading[11] which strongly prefer or even require the subject to
be zero) the pronoun is sometimes used and sometimes not. No
general pattern is readily apparent, and the choice seems to

---

[11]   See also section 4.3. for the possible use of overt
woni in arbitrary plural constructions.

depend on a variety of factors, (including the speaker and the rhythm of the sentence). According to Faßke, of the 190 examples he examined which were not instances of one of the environments above, 75 had a zero subject, while 115 had an overt subject. He thus concludes that it is **not** necessary to prohibit the use of overt pronouns in the literary language categorically ("njetrĕbne je, wuživanje personalnych pronomenow w literarnej rĕči tak kategorisce zakazać").

Jenč (1959) is mainly concerned with the position of the auxiliary and the *l*-participle in Sorbian. However, he also provides some interesting data about the use of personal pronouns in subject position. His investigation is based on a large sample of native speaker recordings. The most important results with respect to our topic are the following. The forms of *być* are quite consistently treated as enclitics, and colloquial speech (*ludowa rĕč*) does not put the auxiliary in initial position except in yes/no questions. The auxiliary is typically preceded by a personal pronoun, as the following examples, taken from Jenč (1959), illustrate:

(9) **ja sym** tam wjele moli byl
I aux$_{1sg}$ there many times was
'I have been there many times'

(10) **wón je** sej tola myslil, zo wón te
he aux$_{3sg}$ refl though thought that he this
šerjene wjedźe
ghost leads
'he however thought that he leads the ghost'

The number of instances where the auxiliary is found in initial position is low. According to Jenč this is mainly the case when people try to make a conscious effort to speak "correctly", and therefore without pronominal subjects. Example (11) below reveals this "extra" effort since virtually all Sorbs normally use German numbers in spontaneous speech.

(11) sym rodźena w Jaseńcy sydomadwaceteho septembra
     dźewjatnaće stow a wósomasydomdźesat
     '(I) was born in Jeßnitz on September 27, 1978'[12]

One individual seemed to use the construction with the auxiliary in initial position intentionally as a stylistic device of terseness, perhaps comparable to the "diary-style" mentioned in Chapter 2.

If something other than the subject is in initial position, the subject pronoun is still usually lexicalized, following the clitic verb form, as shown in (12-14).

(12) tam **sym ja** bydlił
     there aux_{1sg} I lived
     'I lived there'

(13) a nětk **stej wonej** wječor šłoj
     and sometime aux_{2dual} $they_{dual}$ evening went
     'and sometimes they both went in the evening'

(14) za sport **sym ja** wězo přeco zajim měł
     for sports aux_{1sg} I indeed always interest had
     'for sports I surely have always had an interest'

---

[12] This must be a mistake, since the recordings were made before 1959. The speaker probably meant to say 1878.

Jenč points out that the few instances without personal pro-
nouns are produced by people who have jobs in the public sec-
tor and are more likely to try to conform to the school style
in which they were trained. Even those individuals usually
start using personal pronouns more frequently further into the
recording when their style gets more lively and they become
less driven to try to produce "correct" Sorbian.

With respect to subordinate clauses Jenč points out that
there, too, personal pronouns are used more frequently than
they are omitted. In his sample the ratio came out 64% : 36%
in favor of lexicalized personal pronouns. Below are two of
the many examples cited by Jenč:

(15) hdyž sym ja najprjedy služil, da   sym   ja
     when aux$_{1sg}$ I first   served however aux$_{1sg}$ I
     kolosal lochki byl
     very   light was
     'however, I was very light when I first served'

(16) hdyž je   wón wumrěl, to   ja wěm
     when aux$_{3sg}$ he died   that I know
     'I know when he died'

It should be emphasized once more that in Sorbian the use
of overt pronouns is not obligatory, in contrast to German.
Even though the frequent use of overt pronouns in Sorbian is
usually blamed on German influence, there are some examples

where Sorbian speakers omit pronouns in German, which further

supports the claim that pronouns in Sorbian are optional.[13]

There might even be evidence within the literary language

itself that the very limited use of overt pronouns is an arti-

ficial phenomenon, induced by prescriptive grammars and not in

line with the real language. It is well known that clitics in

some languages occupy the second position in the clause, the

so-called Wackernagel position.[14] According to Jenč (1959)

and Michałk (1970) the present, past and conditional parts of

the auxiliary verb *byč* belong to the class of clitics and as

such should not be found in the first position of a clause,

assuming that Sorbian is a Wackernagel language.[15] As we have

seen above in the discussion of Jenč (1959), that expectation is

---

[13]   Michałk (1977) cites the following two examples for
missing personal pronouns in German under the influences of
Sorbian:

(i) # bin mit in die LPG eingetreten (missing *Ich*)
    '(I) have also joined the cooperative'

(ii) Kuhstall ist noch nicht so, wie # sein muß (missing *er*)
     'cowbarn is not yet such as (it) must be'

[14]   If there is more than one clitic, there are rules
for the positioning of those clitics relative to one another.
In Sorbian, according to Michałk (1965b), the relative order-
ing is the following:

    particle/*byč*/dative/accusative/nominative

hač	drje	sym	či	je	ja	tež	prawje	wumyl
if	part.	aux$_{1sg}$	you$_{dat}$	them$_{acc}$	I	also	correctly	washed

'if I indeed washed them the right way for you'

[15]   The status of the forms of *byč* as enclitics is not
absolute. In interrogative clauses the auxiliary is found
regularly in first position both in colloquial speech and
literary style, where it might be analyzed as proclitic.

borne out in normal spoken Sorbian. Typically they are in second position preceded either by a nominal subject NP or by the personal pronoun. However, in the literary language there are many examples with an initial auxiliary.[16] Jenč attributes this phenomenon to the fact that writers feel obliged to comply with the rule of the official grammar that personal pronouns in general must not be used. Example (17) with the conditional form of *być* is taken from Koch (1991), and (18) and (19) are taken from the newspaper *Nowa Doba* as cited by Jenč (1959).

(17) **bych** sej   samomu   zakładne dožiwjenje
     aux$_{1sg\,subj}$ refl$_{dat}$ self$_{dat}$ basic   experience
     pokradył; **bych**   je pódla sebje stejo wostajił
     deprived; aux$_{1sg\,subj}$ it next   myself standing left$_{past}$
     '(I) would deprive myself of my basic experience;
     (I) would just turn my back on it'

(18) **sym** prajił,   zo   doma   wostanjetaj
     aux$_{1sg\,pres}$ said   that at-home will-stay$_{3dual}$
     '(I) said that the two of them will stay at home'

(19) **bě**   nam tež najnowše nowiny sobu přinjesł
     aux$_{3sg\,imperf}$us too newest   paper self brought
     '(he) had also brought the newest papers with him'

Interestingly, even the puristic grammars of the 19th century still describe sentences such as the ones above as ungrammatical. Liebsch (1884: 228), for instance, writes that *sym, sy,*

_____

[16]   This may also be the reason why some linguists, including De Bray (1963), write that auxiliary verbs in Sorbian are not clitics, in contrast to languages like Czech or Serbo-Croatian.

-92-

je, smój, staj, stej, smy, sće, su must never be in the initial position in a sentence.

4.4. Application of null-subject diagnostics

Thus far we have seen that there is a contradiction between the normal spoken language and the exhortation found in various Sorbian grammar books to omit overt subjects. From this it is clear that we should at least be very suspicious about the *pro*-drop status of Sorbian. We will now apply the set of three diagnostics introduced in Chapter 2 in order to independently determine the status of Upper Sorbian with respect to null-subject characteristics. For the reasons discussed in Chapter 3 for Russian, this investigation will focus on the interpretive possibilities associated with *overt* subjects. The three diagnostics are repeated here as (20).

(20) The following holds true for null-subject languages:
   a. **Overt** subjects are stylistically marked.
   b. **Overt** subjects cannot function as bound variables.
   c. **Overt** subjects cannot be arbitrary in reference.

The linguistic data presented in this chapter come from various sources, which should improve reliability. I use examples based on published transcripts of spontaneous speech, examples cited in published articles, Sorbian sentences which I elicited from native speakers on the basis of German source texts, and Sorbian examples constructed by me and judged by native speakers. I conducted most of my fieldwork in the

rural area north-west of Bautzen which was previously referred to as the "Kerngebiet". Since there are only an estimated 2,600 native speakers of Upper Sorbian dispersed outside this area, it was impracticable to undertake an exhaustive search.

With respect to the null-subject criterion in (20a), the rampant use of unemphatic pronouns illustrated in the previous section strongly suggests that Upper Sorbian is not a canonical null-subject language. My fieldwork showed the same high level of overt pronoun use.

A similar conclusion can be drawn from the appearance of the overt *wón* 'he', *wona* 'she' and *woni* 'they' in the bound variable contexts in (21a-i). Example (21) is taken from a literary text and cited by Faßke (1981). Examples (21b-g) were constructed by some informants when they were given the corresponding German sentences. All other informants accepted (21b-g) with a coreferential or bound reading of the pronoun in the subordinate clause when the sentences were given to them in Sorbian. Sentence (21h) was previously listed under (8) to illustrate one of Faßke's (1964) overt pronoun rules. In terms of our diagnostics, Faßke's rule means that a personal pronoun is usually realized phonologically if the subject antecedent of the matrix clause is a full NP, thereby directly contradicting the null-subject criterion (20b). Example (21i) is a sentence which Jenč (1959) uses in the explanation part of his article. Technically speaking, only (21b-d) are true examples of bound variables. However, criterion (20b) also

applies to structures where the QP is replaced by a referen-
tial NP (cf. section 3.1.2).

(21) a. **Feliks** njepytny,   zo **wón** hižo   hodźinu
       Feliks not-realized that he already hours

       po   měsće honja
       through town ran

       'F. didn't realize, that he had already been
       running through town for an hour'

b. **kóždy nan**   mysli, zo **wón** mudreho syna ma
   every father thinks that he smart   son has

   'every father thinks he has a smart son'

c. **kóždy prošer** so nadźija, zo **wón** skoro zbohatnje
   every beggar refl. hopes that he soon become-rich

   'every beggar hopes that he will soon be rich'

d. **kóždy student** sej mysli, zo **wón** jedynku dóstanje
   every student refl thinks that he   A   will-get

   'every student thinks that he will get an A'

e. **Serbja** su   hordźi, zo **woni** swoju rěč   maja
   Sorbs aux$_{3pl}$ proud that they their language have

   'Sorbs are proud that they have their own language'

f. **Jana** je   prajiła, zo chce **wona** rady kupać   hić
   Jana aux$_{3sg}$said   that wants she gladly swimming go

   'Jana said that she would gladly go swimming'

g. **Jana** lubuje toho muža, na kotrehož je   so **wona**
   Jana loves   that man   at whom   aux$_{3sg}$ refl she

   před dźesać lětami wudała
   ago ten   years married

   'Jana loves the man whom she married ten years ago'

h. a **mój nan**   je   prajił, zo bdźe **wón** přećiwo
   and my father aux$_{3sg}$ said   that will he   against

   Vogtej hić
   overseer go

   'and my father said that he would go against the
   overseer'

i. Ale tež **Grofa** staji, hdyž **wón** žiwišo rěči,
but too Grofa puts   when he lively speaks
personalny pronomen.
personal   pronoun

'but also Grofa uses, when he speaks lively,
a personal pronoun'

The application of the third null-subject criterion
produces the same results, although perhaps less obviously so.
While I did not observe the use of overt *woni* in spontaneous
speech, there were some informants who used this construction
when translating German sentences with arbitrary reference
into Sorbian. However, the main point is that most infor-
mants, especially those without literary training, agreed that
the presence of the overt 3rd person plural pronoun does not
force a referential reading of the pronoun, in contrast to
"true" null-subject languages such as Czech. Examples of
overt pronouns with arbitrary reference are given below.

(22) a. w Londonje su   **woni** zaso stawkowali
in London aux$_{3pl}$ they again struck
'in London they went on strike again'

b. we Lužicy **woni** maja rjane stare naložki
in Lusatia they have nice old   customs
'in Lusatia they have nice old customs'

c. nětko **woni** wšudźe   wo přichodźe Serbow
now   they everywhere on future   Sorbs$_{gen}$
diskutuja
discuss
'nowadays they are talking everywhere about the
future of the Sorbs'

d. někdy su woni skromnišo žiwi byli hač dźensa
   once aux$_{3pl}$ they modestly living were than today
   'once they used to live more modestly than today'

e. na tutej kupje **woni** hižo 500 lět francosce
   on this island they already 500 years French
   rěča
   speak
   'they have been speaking French on this island
   for 500 years'

f. na wikach **woni** sad předawaja
   at markets they fruit sell
   'they are selling fruit at the market'

g. w Sewjernej Americe **woni** jendźelsce rěča
   in North America they English speak
   'they speak English in North America'

I therefore conclude that Upper Sorbian is **not** a canonical null-subject language. This raises two main questions:

A. Why do the authors of **most grammars** try to describe Sorbian as a null-subject language (in contrast to the actual spoken language) by strongly discouraging, if not even prohibiting, the use of overt non-emphatic pronouns?

B. Why is Sorbian different from the other West Slavic (and South Slavic) languages?

The next section is devoted to the first question, while section 4.6. will focus on question B.

4.5. Pronouns and purism

I agree with Faska (1964), who assumes that the pronoun-pro-
hibiting rules of the traditional grammars were actually not
formulated on the basis of the Sorbian language, but instead
under the influence of Czech. Since the literary norm is
based on the linguistic works of the middle and late 19th
century, it is obvious how the pronoun prohibition rule made
its way into current school grammars as well.

This seemingly small detail of how to use overt pronouns
can be seen as reflecting a much larger phenomenon when viewed
in historical context. In the mid-19th century there was a
strong Sorbian national movement known as the *serbske narodne
wozrodźenje* ('Sorbian National Rebirth'), which led to a wide
application of the literary language to secular situations.
The Sorbs understood themselves to be part of the larger Slavic
family of peoples, in keeping with the "Panslavism" of this
period, and looked especially to their Czech neighbors as a
guide. This was accompanied by the phenomenon of "purism",
aimed at complete removal of all foreign, and especially German,
elements from the Sorbian literary language. New lexical
needs were satisfied by the introduction of loan-words and
loan-translations from other Slavic languages, especially
Czech. Similar attempts were aimed at a removal of certain
"non-Sorbian" morphological and syntactic phenomena from the
literary language, most of which were attributed to German
influence.

One example of these "alien" elements, which can be used to date the rise of purism fairly precisely (within 10 years), was the use of the demonstrative pronouns *tón, ta, te* in the same distribution as the definite article in German. As Lötzsch (1968) points out, analyzing different descriptions of the use of the definite article in Sorbian, Seiler's (1830) grammar is still free from puristic biases. Seiler tries to give an objective analysis of the use of the demonstrative pronoun as the definite article in Sorbian. Jordan's (1841) grammar, on the other hand, may be the first puristic work of Sorbian linguistics. He refers to the use of the definite article in Sorbian as "Abnormalität" and "grundlose und wider-sinnige Nachahmung des deutschen Idioms, die dem Charakter und dem Geiste der slawischen Sprachen rein zuwider läuft" (Jordan 1841: 138). Schneider (1858: 8) calls the definite article in Sorbian a "Danaergeschenk der Deutschen".[17]

Similarly, the use of overt pronouns in Sorbian has often lbeen blamed on German influence, although this is not neces-sarily the case. We will return to the question of why Sorbian

---

[17]   The use of demonstratives as definite articles is still discouraged for the literary language but widely spread in spoken Sorbian. The following sentences are examples from colloquial Sorbian, taken from Michalk and Protze (1974).

(i)   ja sym slyšal, zo jo wón sej tu nohu zlamal
       ich habe gehört, daß er sich das Bein gebrochen hat

       'I heard he broke his leg'

(ii)   te ludźo su tola wšitcy wonka na naměsće
       die Leute sind alle draußen auf dem Anger
       'the people are all outside on the village green'

is not a *pro*-drop language in more detail below. Even if the use of overt pronouns proves to be genuinely Sorbian, language purism would have influenced this phenomenon because it **appeared** to be of German origin, simply because it was different from the *pro*-drop Czech (and Polish) neighbors. This turning to other Slavic languages to cleanse their own was a characteristic trait of the Sorbian purists.

Language purism in general is very prescriptive in character, and it contributed to a gap between the literary language and spoken language in Sorbian. This gap has only widened, since this crusade did not stop in the past century, unlike other puristic movements. Jenč (1959: 43) refers to this when he accuses contemporary writers of clinging to grammatical norms of other Slavic languages as if "hypnozowani" ('hypnotized'). According to Jenč (1964: 1965), the difference between written and spoken Sorbian today is such that one might rather speak of two different languages instead of two different forms of the same language. He points out that there are only a few differences between the colloquial language today and the written language before the middle of the 19th century. We can thus assume that formerly there was no difference between spoken and written language and that it was (and still is) the search for an "ideal" and "pure" Sorbian which has been moving them further apart by prescribing archaic and artificial norms for the written language. In several articles Jenč appeals to the Sorbian writers of today's literature

to overcome this schism and stop their "wěčne bojazliwe hladanje na druhe rěče" ('constant fearful look at other languages'). As Luther would have said, "man muß dem gemeinen Mann aufs Maul schauen" ('one should look at the common man's mouth') to codify a language so that it can function as a widespread means of communication.

The results of purism with respect to the use of personal pronouns in the written language can be illustrated by the following Biblical passage as it appeared in 1728 and 1896 versions, as cited by Jenč (1964: 252).

(23) a. Matthew 25:42,43 (1728, Richter (ed.))

Přetož ja sym hłódny byl a wy njejsće mi jěsć dali. Mi je so pić chcyło a wy njejsće mi pić dali. Ja sym hósč byl a wy mje njejsće hospodowali; ja sym nahi byl, a wy mje njejsće wodźěwali; ja sym chory a jaty byl, a wy njejsće ke mni pohladali.

b. Matthew 25:42,43 (1896, Lusčanski and Hórnik)

Přetož buch hłódny, a njedašće mi jěsć; buch lačny a njenapojišće mje. Běch cuzy, a njehospodowašće mje; nahi a njewoblekašće mje; khory a w jastwje a njewopytašće mje.

There are nine overt personal pronouns in the version of this passage translated in 1728, none of which are present in Lusčanski and Hórnik's 1896 version in (23b). We have to assume that this change happened under puristic pressure since there are no indications that Sorbian had really become pro-drop (again).

Michaĺk (1972: 91,92) gives some interesting numbers with respect to the sudden decline of the use of overt pronouns in successive versions of Bible translations. The various Catholic translations seem particularly relevant since these are all based on the Latin Vulgate, i.e. direct influence from German can be excluded. Swětlik (*Perikopen*[18] 1690) uses 20 personal pronouns in Matthew 2:2-18 which are not present in the Latin. In the 1750 pericopes, which differ only slightly from Swětlik's version, there are still 17 such pronouns. In the pericopes from 1848 this number is reduced to 7, in Buk's (1862) translation there are only 3 left, and they are all stricken from Lusčanski and Hórnik's (1896) version.[19]

In the 1896 version we find that there are personal pronouns **only** in those places where they correspond to personal pronouns (1st and 2nd person), demonstratives or *qui* in the Latin version. Below are some examples:

---

[18] Pericopes are selections of various Bible passages.

[19] The Protestants translated their version of the Bible from Luther's German text. Interestingly enough, the number of non-emphatic personal pronouns is actually lower than in the Swětlik translation. In Frencel (1670) there are 13 personal pronouns in Matthew 2:2-18 which have no correspondence in the Hórnik/Lusčanski version. This number stays constant in the 1728 and 1823 versions, though in slightly different distribution. Obviously, the Protestants were not as concerned about "purifying" their Bible, because the 1905 translation still has the same 13 personal pronouns. It is not until 1960 that we note a drastic drop in this number from 13 to just 1, apparently due to a post-1945 puristic movement.

-102-

(24) Matthew 3:11

    ego quidem baptizmo vos in aqua ..
    ja drje kŕću was z wodu

(25) Matthew 2:13

    qui cum recessissent ..
    Hdyž běchu **woni** wotešli

(26) Matthew 2:5

    at **illi** dixerunt ei
    woni pak prajachu jemu

We have already seen Jordan's account of overt subject pronoun use in Sorbian in which personal pronouns are only used in emphatic contexts, as is typical for Czech. Only ten years earlier, however, Seiler states the complete opposite:

> Die persönlichen Fürwörter werden zum Zeitworte **erfordert** ... (Seiler 1830: 113).[20]

He then lists rules by which personal pronouns are omitted:

> Gebraucht und wiederholt werden sie nicht: a) wenn das Subjekt oder sein Pronomen im vorhergehenden Satze .... genannt worden ist ... b) im Imperativ und in Gemeinplätzen, wo man eine bestimmte Person geradezu nicht meint. (ibid)[21]

---

[20]   'The personal pronouns are **required** with the verb.'

[21]   'They [personal pronouns, ML] are not used and repeated: a) if the subject or its pronoun ... was mentioned ... in the previous sentence ... b) in the imperative and in general meaning where one does not refer to a certain person.'

It seems that Seiler's description is certainly closer to the real situation found in spoken Sorbian today. It also gives a more realistic account of the use of pronouns in the Sorbian language of his time. For instance, in the preface to his new political newspaper *Serbski powĕdar a kurĕr*, first published in 1809, J. B. Dejka uses 10 (non-emphatic) overt subject pronouns (5 x *ja*; 2 x *wy*; 3 x *my*) and only uses a zero pronoun twice.[22]

The influence of the Czech language on the repudiation of overt pronouns in Sorbian is even more likely when one takes into consideration the fact that three of the authors of the aforementioned grammars studied and worked in Prague at the *Serbski seminar* ('Sorbian Seminary') at the time they wrote their grammars, i.e. Jordan (1831-38), Liebsch (1880-83) and Kral (1879-89). Lusčanski, the co-author of the 1896 Bible, was head of the *Serbski seminar* in Prague from 1877 to 1894.

4.6. Why is Sorbian not a canonical null-subject language?

The conclusion that Upper Sorbian is not a *pro*-drop language in the classic sense raises the question of **why** Sorbian is different from canonical null-subject languages, such as Czech and Polish. As we have seen in (1), the verbal paradigm of Sorbian is morphologically uniform, and null subjects should

---

[22]   In Lorenz (1981: 106).

thus be **licensed.**[23]   So from the perspective of a licens-
ing/identification approach to the problem the following
question presents itself: why is the **identification** of refer-
ential subjects in Sorbian blocked and Sorbian thus prevented
from being a canonical null-subject language?   Two possible
analyses are presented below.   These are then followed by a
discussion of the limitations of a "pure" licensing/identifi-
cation approach to the problem and a presentation of a third,
alternative analysis.

4.6.1. Identification and the number of syncretisms

A possible explanation for why Sorbian is not a null-subject
language might be formulated in terms of Roberts's (1993)
restriction on the number of syncretisms tolerable for suc-
cessful identification.

Roberts adopts Rizzi's (1986) licensing criterion for
null subjects:

(27) If X^0 is the licensing head of *pro*, X^0 can assign
Case to the position occupied by *pro*.

As pointed out by Roberts (1993: 18), it has been proposed that
structural case can be assigned in two different configura-
tions, government or agreement, as illustrated below.

---

[23]   The issue of overt, non-referential pronouns will
be discussed in Chapter 5.
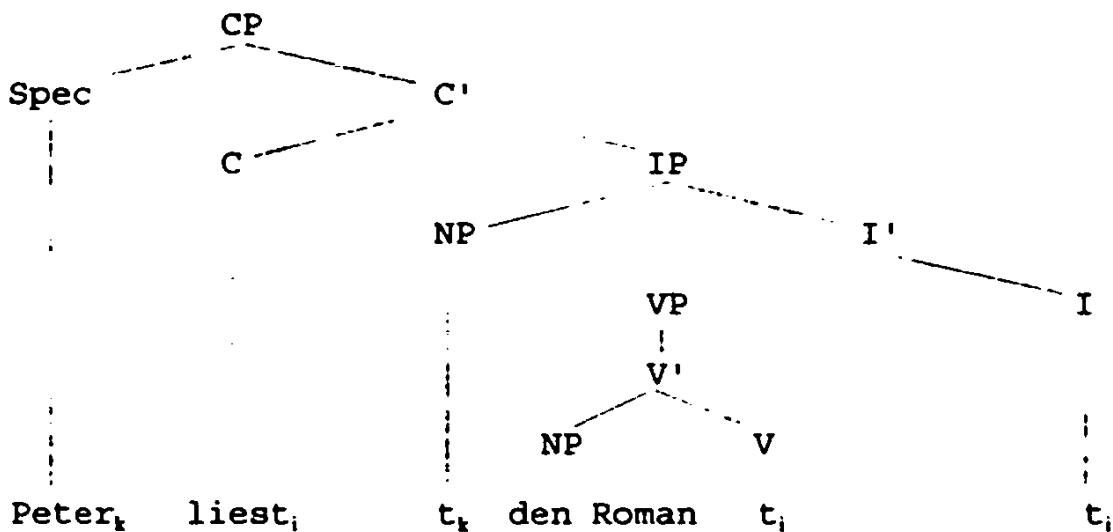
(28) Case assignment under government:

a. X'
```
     X'
    /  \
   X    NP
```

b. X'
```
      X'
     /  \
   X⁰    YP
        /  \
      NP    Y'
             |
             Y
```
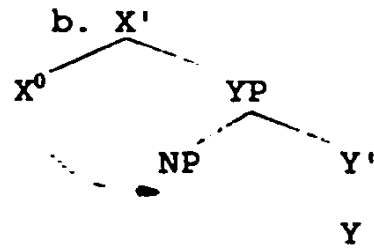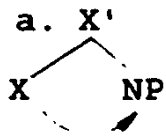
(29) Case assignment under agreement:

```
      XP
     /  \
   NP    X'
          |
          X⁰
```
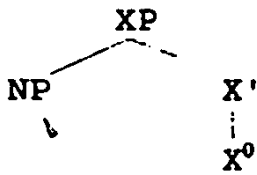
There is assumed to be a nominative assignment parameter, according to which nominative can be assigned under government or under agreement. Nominative assignment under government presupposes movement of the inflected Verb into C position to produce the configuration shown in (28b). This is typical for V-2 languages, such as German, as illustrated in (30).[24]

(30) Peter liest den Roman. 'Peter reads the novel'

```
                    CP
                  /    \
              Spec      C'
               |       /  \
               |      C    IP
               |         /    \
               |        NP     I'
               |         |    /  \
               |        VP    I
               |         |
               |        V'
               |       /  \
               |     NP    V
            Peterₖ  liestᵢ  tₖ den Roman  tᵢ        tᵢ
```

Nominative assignment under agreement, on the other hand, is normal spec-head agreement between AGR as the head and *pro* in the specifier position of AgrP.

The identification of referential *pro* (which Roberts calls "content licensing") can happen in two different ways: by "functional richness" and by "formal richness". The precondition on formal richness, according to Roberts, is a morphologically uniform paradigm where an agreement slot is available throughout.[25] Functional richness, on the other hand, is defined in terms of a restriction on the number of syncretisms tolerable for successful identification, and it allows for one syncretism in the paradigm. Functional richness, as opposed to formal richness, is also compatible with a zero ending (and since the zero ending can be the syncretism, there is actually the possibility of two zero endings in a functionally rich paradigm as long as all the other endings are distinct).

We have discussed above how the Sorbian paradigm, and in fact the verbal paradigms of all Slavic languages, fulfill the morphological uniformity condition, which is treated as the licensing condition within the Jaeggli and Safir system, as opposed to Roberts, for whom it is one of the two possible conditions for **identification**. If we want Roberts's system to

---

[25]   In contrast to Jaeggli and Safir's morphological uniformity hypotheses, Roberts's formal richness pertains only to paradigms with consistently derived inflectional forms and not to languages which do not have agreement morphology.

work we must find a reason why formal richness, i.e. morpho-
logical uniformity, is not relevant for identification in
Slavic.

It is crucial that a language may not automatically avail
itself of either of the two mechanisms of identification. In
this context, Roberts (1993: 207) proposes the following gen-
eral principle, (mainly to account for the loss of preverbal
pro in French after a change in the nominative assignment
parameter as summarized below).

> (31) If Agr^0 formally licenses *pro* only under agreement,
> then Agr^0 must be functionally rich.

Roberts's theory is based on the history of French, which
once had null subjects, even though *pro* was only found in cer-
tain syntactic environments, unlike in modern Polish or Ital-
ian, for example. He shows how the loss of null subjects in
French can be linked to the change of a single parameter,
namely the nominative case assignment parameter, in conjunc-
tion with principle (31). Old French was a null-subject lan-
guage in the sense that referential null subjects were possi-
ble in V-2 contexts, i.e. postverbally under government by
AGR, which also means that nominative was assigned under gov-
ernment. Identification in early Old French was possible
because the verbal paradigm was functionally rich. A major
restructuring took place in the 12th century, and as a result
the paradigm ceased to be functionally rich due to the large

number of syncretisms. However, at the same time the verbal paradigm became formally rich instead. Thus, despite the loss of functional richness, French remained a null-subject language for a time, as long as nominative assignment under government was available. Middle French saw a change with respect to the occurrence of null subjects in that the number of contexts in which null subjects were allowed actually increased -- specifically null subjects could also occur preverbally. According to Roberts this was possible because nominative assignment was no longer restricted to government configurations, but instead expanded to include the agreement context as well. Middle French had a morphologically uniform paradigm and content licensing was accomplished via formal richness.

The relatively sudden disappearance of null subjects in the 17th century is attributed by Roberts to the fact that nominative assignment became restricted to the agreement configuration, i.e. nominative cannot be assigned under government. At this point, post-verbal null subjects are directly excluded since they cannot be licensed anymore. What about pre-verbal null subjects though? According to Roberts, as a consequence of the change in the nominative assignment parameter, null subjects in a language like 17th-century French can only be identified by a functionally rich paradigm, following principle (31). Formal richness does not suffice any more.

However, the verbal paradigm of French does not fulfill the functional richness criterion and identification fails.[26]

In order for Roberts's approach to be at all applicable to identification in Sorbian we have to make the assumption that the nominative assignment parameter is set to "agreement only".[27] In other words *pro*, is formally licensed under agreement only, and for referential subjects to be identified we need a functionally rich paradigm, while formal richness (i.e. morphological uniformity) is not applicable. A tolerance of one syncretism for functional richness, as defined by Roberts, may indeed be necessary to incorporate some Slavic null-subject languages. For example, there is a syncretism in Bulgarian between 2nd and 3rd person singular endings in both the imperfect and the aorist, as illustrated in (32) and (33).

---

[26] Sprouse and Vance (1994) point out a rather counterintuitive point of Roberts's analysis. Middle French was a language which had nominative assignment both by agreement and by government. However, assuming the validity of principle (31) and taking into account that the paradigm was only formally rich, null subjects in the agreement configuration could only be licensed because of the instances where *pro* was licensed in the government configuration.

[27] This assumption is far from universally held. At first glance one might try an analysis, according to which Slavic languages have different settings for the nominative assignment parameter and thus different ways of identification available to them. However, the setting of the nominative assignment parameter makes no difference for the canonical null-subject languages, since they all seem to have both formally and functionally rich paradigms. For Russian and Sorbian, on the other hand, the setting of this parameter is crucial. King (1993) proposes nominative assignment under government for Russian. However, this would make formal richness relevant for identification, with the undesirable result that Russian should be a canonical null-subject language.

(32) jadeše   Imperfect: 'eat' 2nd/3rd sg

(33) jade   Aorist: 'eat' 2nd/3rd sg

Note that this situation does not count as two syncretisms, presumably, because they occur in different tenses. To accommodate these data we would need to revise Roberts's syncretism criterion slightly:

(34) A verbal paradigm is functionally rich if there is
     no more than one syncretism per tense.

Somewhat more problematic might be the situation in Old Church Slavonic, where one finds two syncretisms within a single tense. In OCS both the imperfect and the aorist display, in addition to the syncretism just mentioned for Bulgarian, a syncretism between the 2nd person plural and 3rd person dual, as illustrated in (36) and (38).

(35) jaděaše   Imperfect: 'eat' 2nd/3rd sg

(36) jaděašete Imperfect: 'eat' 2nd pl or 3rd dual

(37) jastъ   Aorist: 'eat' 2nd/3rd sg

(38) jaste   Aorist: 'eat' 2nd pl or 3rd dual

It can be shown (cf. Lindseth and Franks 1994; Franks 1994) that the use of the same form for the 2nd person plural and 3rd person dual does not count as a syncretism in the relevant sense. A syncretism reflects a single form that is non-distinct from more than one feature complex, i.e. it is a defective feature matrix. Verb forms that are lexically associated

-111-

with such defective feature matrices can be substituted for multiple, fully specified AGR matrices at S-structure. A form, such as (35) *jaděaše* '(you-sg/(s)he) was eating', for instance, can be analyzed as having the pronominal features [-1st,+sg], rendering it non-distinct from both 2nd and 3rd person singular agreement. However, when the features cut across both person and number, as in (36) *jaděašete* '(you-pl/(s)he-dual) was eating', this is **not** an instance of feature neutralization, because it cannot be expressed using a single entry in the lexicon. The fact that this is technically not a syncretism provides a possible explanation for why identification of null subjects is not blocked in OCS.

Now compare the Upper Sorbian forms below with the OCS forms in (35-38):

(39) piješe   Imperfect: 'drink' 2nd/3rd sg

(40) piještaj   Imperfect: 'drink' 2nd/3rd dual

(41) wupi   Aorist: 'drink' 2nd/3rd sg

(42) wupištaj   Aorist: 'drink' 2nd/3rd dual

In contrast to OCS, Upper Sorbian exhibits two real syncretisms within a single tense, violating condition (31) for the identification by functional richness. Common Slavic had distinct forms for the 2nd and 3rd person dual. In (Upper) Sorbian there is no distinction between the 2nd and 3rd person dual (in addition to the inherited syncretism between the 2nd and 3rd person singular in the imperfect and aorist). The two

different forms of the Common Slavic 2nd and 3rd person dual have been reinterpreted as a gender distinction, at least as far as the literary language is concerned.

Table 3: The Dual in OCS and in Upper Sorbian

		OCS	Upper Sorbian
Dual Present	**1st**	neseve̊	njesemoj
	2nd	neseta	njesetaj/njesetej
	3rd	nesete	njesetaj/njesetej

In literary Upper Sorbian the -*taj* ending is used in reference to male persons, while -*tej* is used for all other genders. However, most Upper Sorbian dialects do not express this distinction and use only the ending -*tej*.

We have seen that an approach to identification in terms of Roberts's restrictions on the number of allowable syncretisms seems to provide an account for the lack of identification in Upper Sorbian.[28] However, there is one major drawback to this sort of morphological explanation: This account makes the bizarre prediction that the loss of the dual in Upper Sorbian should reinstate the successful identification of null subjects in Upper Sorbian. The Sorbian verbal paradigm without the dual would be "functionally rich" in the

---

[28] Roberts's account is intended to explain why French lost null subjects entirely. A strict interpretation of Roberts would thus predict that Sorbian has no null subjects at all, and could not account for the apparent optionality of subject pronouns. In this context see also 4.6.3. for a discussion of the general challenges the Sorbian situation provides for identification/licensing frameworks, and 4.6.4. for an alternative analysis.

sense defined by Roberts and thus no violation of principle (31) would occur.

Ideally, we should compare data from Old Sorbian before and after the merger of the 2nd and 3rd person dual to see whether this change brought about a concomitant increase in the use of overt subject pronouns. Unfortunately, this is not possible, because the earliest Sorbian texts are relatively recent (16th century) and already show an extensive use of overt personal pronouns in subject position, which is similar to today's colloquial language. There might be yet another way to determine the validity of Roberts's syncretism account for Sorbian. Some dialects of Upper Sorbian have only infrequent use of the dual number. Fieldwork might be able to determine whether or not there are dialects which lost the dual completely, and whether or not such dialects really have the relevant null-subject characteristics which were found to be missing in the dialects which preserved the dual.

In this context, let us return to Russian. On the basis of the null-subject criteria I argued that Russian is not a canonical *pro*-drop language. An explanation for the lack of identification of referential null pronouns was given in terms of the lack of AGR$_{pers}$ in the past tense. We might now try to see whether Roberts's theory can be applied to Russian as well. It turns out that "defects" in the past tense paradigm can once again be interpreted as the culprit for blocking identification. Russian past tense verbs are marked for gen-

der in the singular, but this opposition is completely neu-
tralized in the plural.

Table 4: Russian Past Tense

Sg.	**Masc.**	byl	**Pl.**	byli
	Fem.	byla		
	Neuter	bylo		

Since the singular distinguishes masculine, feminine and
neuter, there must be a double syncretism of gender features
in the plural. Identification by functional richness thereby
fails.

4.6.2. A word order account

While it is in principle possible to formulate an explanation
for why Upper Sorbian is not a null-subject language in terms
of a restriction on the number of syncretisms in the morpho-
logical marking, I also pointed out the counterintuitive pre-
dictions associated with this approach.

The non-*pro*-drop properties of Sorbian may also be at-
tributable to some other aspect of its syntax. An inspection
of Sorbian texts reveals a striking preponderance of verb-
final declarative sentences, in contrast to other Slavic lan-
guages, such as Czech, for example. The sentences below are
taken from Michałk (1962b) and are drawn from both literary
and dialectal sources.[29] Sentences (43) and (44) are examples

---

[29]   Some of these sentences contain non-standard forms.

of verb-final subordinate clauses; (45-49) illustrate the
verb-final position in main clauses.

(43) nan   sebi mysleše, zo   so   na druhi dźeń
     father self thought that refl at other day

     z   našim susodom **wróći**
     with our   neighbor will-return

     'Father thought to himself that he would return
     with our neighbor on the next day'

(44) Marka so woprawdźe wjeseli, zo   jónu skoro zaso
     Marka refl really   rejoices that once soon again

     do Lipska **pojědźe**
     to Leipzig will-drive

     'Marka is really glad, that she **will** soon once
     again go to Leipzig'

(45) a   hižo   psaj   kaž   zło njedobro přez
     and already dogs$_{dual}$ like evil not-good through

     rěpu   na Jančowu **čěrještaj**
     beets at Jančowa ran$_{dual}$

     'and already the two dogs came running like the
     devil through the beets towards Jančowa'

(46) tutón   jeho pak   na puć k šibjeńcy **dowjedźe**
     that$_{masc}$ him$_{acc}$ however at way to gallows will-lead

     'but that one will lead him on the way to the gallows'

(47) ta   husto po Marku na spěwanje **přińdźe**
     that$_{fem}$ often for Marka at singing comes

     'she often picks up Marka to go singing'

(48) mój   z Marku najwjacy na polu **stojimoj**
     we$_{dual}$ with Mark most-often at field stand$_{dual}$

     'I and Marka stand the most on the field'

(49) najradšo   so   ludźo njedźelu **wobžlokaja**[30]
     most-gladly refl people Sunday   get-drunk
     'people prefer to get drunk on Sunday'

In other Slavic languages the sentence-final position
may, of course, also be occupied by the finite verb, especial-
ly if the verb is the rheme, i.e. the element which provides
the newest, most important part of information. In general,
the final position in those languages seems to be reserved for
the rheme. However, in the Sorbian examples above consider-
ations of functional sentence perspective should not have
contributed to the final position of the verbs. What is im-
portant for our discussion is that the **unmarked** position for
the finite verb in both the literary and colloquial language
seems to be clause-final. Michałk (1962a; 1962b) provides
some interesting figures. The ratio of verb-final versus
verb-internal position came out to be 62% : 38%. According to
Michałk, in 55% of all analyzed sentences with the finite verb
in final position the verb did not belong to the rheme. In
86% of all cases where the finite verb did not show up in
final position, the verb was immediately followed by the rheme
in final position. This indicates that the verb can be dis-
placed from its unmarked final position in order to put the
rheme there instead. The tendency to put the finite verb into
final position, even if it is not the core of the new informa-

---

[30]   Most likely intended to be *wožlokaja*.

tion, was found to be even stronger in the literary language than in colloquial speech.

As far as subordinate clauses are concerned, this structure is, of course, very reminiscent of German, where the finite verb must always be in final position.[31] In Sorbian this verb-final tendency is true for both subordinate clauses and main clauses, though it is just a preference and not an ironclad rule. In German, the finite verb must be in the second position in main clauses. The Sorbian verb can occur in non-final position, especially when some element other than the verb is clearly the rheme.

Based on this discussion I propose that the underlying word order in Sorbian is SOV. For German, SOV order is often derived by analyzing AGRP (and VP) as right-headed, so that when V raises to AGR it will end up in clause-final position, as shown in the structure (50) below:

---

[31] Even in subordinate clauses, Sorbian word order is not identical to that of German. In the perfect and past perfect, the auxiliary typically follows in second position right after the conjunction (in 87% of all cases according Michałk 1970: 23), while the *l*-participle is in final position. Compare the Sorbian and German examples below.

(i) hdyž sym₁ ja z chórom w B. była₂

(ii) wenn ich mit dem Chor in B. gewesen₂ bin₁
'when I have been in B. with the choir'

-118-

(50)   AGRP

    NP   AGR'

          VP   AGR

      NP   V

    S   O t_i V_i
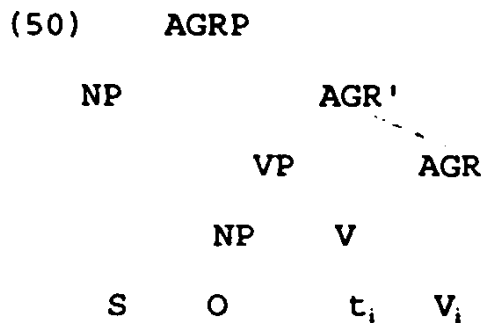
I propose that the right-headedness of AGRP also accounts for word order in Sorbian, the difference being that since AGR (unless it is a clitic) does not raise to C in Upper Sorbian, conjugated verbs are final in both main and embedded clauses.

In analytical tenses, such as the perfect and past perfect, the auxiliary is typically in second position, while the *l*-participle occupies the final position. Consider the following examples, taken again from Michałk (1962b):

(51) ja **sym** někotru zakrasnjenu myš **popanydła**[32]
    I aux_{1sg} some   nice   mouse caught
    'I caught some nice mice'

(52) mój nan   **bě**   tydźeń prjedy wotročka
    my father $aux_{3sg\ imperf}$ week   ago   farmhand
    naswaril
    scolded
    'my father had scolded the farmhand a week ago'

(53) mi so zda, zo **sym** prawje wulku rybu **popanyl**
    me refl seems that aux_{1sg} really big   fish caught
    'it seems to me that I caught a really big fish'

---

[32]   This is very reminiscent of German matric clauses:
    Ich **habe** einige Mäuse **gefangen**.
    'I've caught some mice'

Referring back to the proposal in Chapter 3, I assume the existence of different AGR projections (AGR$_{pers}$ and AGR$_{gend}$) in Slavic. The examples above seem to indicate that all functional projections of the verb except for CP are right-headed. This is illustrated in the structure in (62):[33]

(54)　　　CP
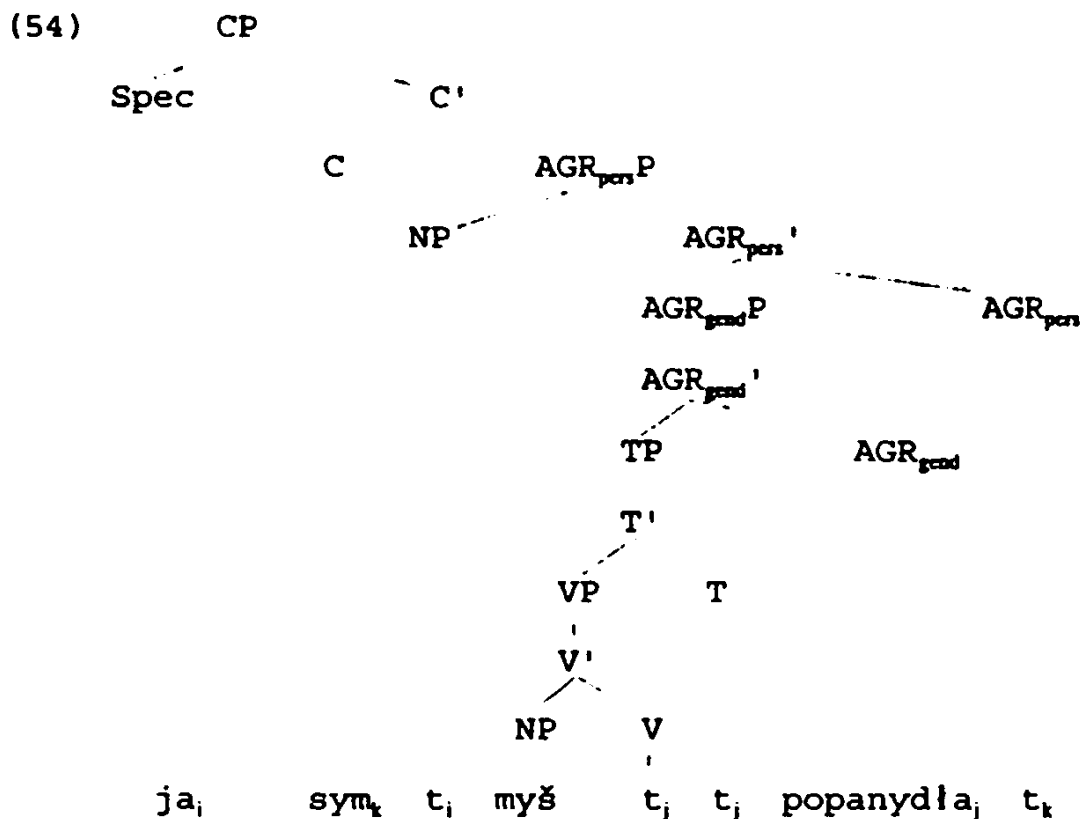
It is thus possible that the lack of identification of null subjects in Sorbian is somehow connected to the direction

---

[33]　　If the auxiliary is negated, it loses its clitic status and is found most often in final position (cf. Jenč 1959: 27-28), as illustrated in the example below. Presumably it then occupies the head position of AGR$_{pers}$ or NegP.

　　　　ja pak　　jeho widźał njejsym
　　　　I however him　saw　　not-aux$_{1sg}$

　　　　'I, however, didn't see him'

of branching of AGR_{pcn} (AGR_S). I speculate that the problem in Upper Sorbian may have to do with the non-adjacency of the head of AGRP ($AGR_{pcn}P$), (which would be the AGR carrying the pronominal features necessary for identification), and the specifier, (which would be the position for *pro*). This non-adjacency, which makes AGR, despite its richness, somehow not accessible to the null subject, is underlying in all sentences, and is also usually present at S-structure, unless a clitic is present as shown in structure (54).[34] It is possible that Sorbian developed its predilection for the final position of the finite verb under the influence of German subordinate clauses. However, Sorbian has generalized that structure in its own way, and the problem for the identification of referential subject pronouns can thus be found within the Sorbian syntax itself.

Just as with the syncretism approach to the issue of null subjects (or lack thereof) in Sorbian, there are also some potential problems for the word-order based account. The literary language is just as verb-final as the colloquial language, if not even more consistently so. However, we have also seen that in literary Sorbian, overt subjects are usually omitted. This situation raises the question of how these

---

[34] Note that German, despite its uniform realization of a relatively rich AGR_{pcn}, is also not a canonical null-subject language. One might speculate that in German too, the fact that thematic null subjects cannot be identified has something to do with the non-adjacency of the head of AGRP and the specifier position (but cf. also fn. 14, Chapter 2).

frequently occurring null subjects are identified in literary Sorbian if identification is blocked by non-adjacency of AGR_{pers} and the subject in $[Spec,AGR_{pers}P]$. It is also the case that there are many instances of null subjects in colloquial Sorbian. If one denies the possibility of identification these should be treated as discourse-licensed ellipses. That would imply that null subjects in literary Sorbian and in colloquial Sorbian are instances of different phenomena, which does not seem very likely. Finally, as pointed out in Lindseth and Franks (1994), one obvious problem with implicating the direction of branching in blocking identification is that there are SOV null-subject languages, which raises the question of why adjacency should be a factor in Sorbian or German, but not in Turkish or Hindu.

4.6.3. Challenges for the identification/licensing framework

The problems mentioned in the previous sections arise when we attempt to solve the issue of *pro*-drop in terms of a highly restricted framework of syntax, trying to account for linguistic phenomena in terms of setting binary parameters and formalizing precise conditions for the occurrence of certain elements. The paradox is that any non-linguist is able to see **why** it is possible to omit subject pronouns in both literary and colloquial Sorbian: They are redundant because the verbal endings carry overtly the information which is necessary to recover the pronominal content of the missing pronouns.

-122-

In general, it seems that most versions of *pro* licensing/identification, such as Rizzi (1986), Jaeggli and Safir (1989), Rohrbacher (1995), Speas (1994), and also the ones discussed in the previous two sections (Roberts (1993) and the word-order account) will encounter two major challenges with respect to the Sorbian data.[35] First, such an approach cannot account for the optionality of overt and null-pronominal subjects in Sorbian. Most versions of the licensing/identification framework predict that a language will either license/ identify null subjects or it will not. Second and related to the first problem, they cannot predict the difference in frequency of null or overt pronouns with different persons of the verb, especially if such a frequency difference cannot be attributed to functional compensation for missing morphological marking on the verb.

We have seen that literary Sorbian does comply with the "subject-pronoun-prohibition rule" given in puristic grammars,

---

[35] An analysis à la Speas fails to account for the whole range of Sorbian data, although at first glance it might appear to be wide enough to cover systems like Sorbian (though certainly not in a very interesting way). Speas considers "strong" agreement as a necessary, but not sufficient condition to treat agreement affixes as individual lexical items. Sorbian fulfills the conditions on "strong" agreement, and should thus have the option of having lexically listed affixes. Under Speas's analysis, Sorbian has chosen otherwise, by listing agreement affixes as part of the verbal paradigm. However, this should make the occurence of any null subjects a violation of the principle of economy, and we cannot account for the optional character of null subjects. Moreover, this type of analysis would also fail to give an explanation for why Sorbian is different, and instead blame it all on a "lexical idiosyncracy" of the Sorbian agreement morphemes.

especially with respect to 1st and 2nd person subject pro-
nouns. However, nonemphatic 3rd person pronouns do occur more
frequently in literary Sorbian.[36] The following are examples
of non-emphatic pronoun subjects from the novel by Koch (1991)
which was mentioned above for its zero subjects:

(55) a. Rěč   zetkanja zmóžni.   **Wona** tež
language meetings will-enable it   too
powitanje přewozmje
greeting will-take-over
'The language will make meetings possible. It
will also serve to greet.'

b. Spisowaćeljo, kotřiž so   w přidatnej rěči
writers   who   refl in another   language

wuznaja su   w přidatnym swěće žiwi. **Woni** su
know aux_{3pl} in another   world live they aux_{3pl}

pola susoda   doma.
at   neighbor at home

'Writers who know an additional language,
live in an additional world. **They** are at home
at their neighbor's.'

Also recall the literary example (21a) repeated here as (56),
which shows that even in the literary language a personal
pronoun may be coreferential with the matrix subject.

(56) **Feliks** njepytny, zo **wón** hižo hodźinu po měsče honja
'F. didn't realize, that **he** had already been running
through town for an hour'

A particularly striking (and *apropos*) example is the following text found in the Šewc grammar (1974: 118). Bear in mind that this is the same grammar which states that there is only a very limited use of personal pronouns in Upper Sorbian.[37]

> Personalne pronomeny *wón, wona* ... njepoćahuja so ... jenož na wosoby, ale tež na zwěrjata a wěcy. **Wone** maja přez to zdobom šěrši woznam. Wosebitosć personalnych pronomenow 3. wos. zwuraznja so dale w tym, zo poćahuja so **wone** zwjetša na něšto, štož je hižo z konteksta znate. **Wone** wobhladuja so tohodla jako wosebita skupina pronomenow ...

> 'The personal pronouns *wón, wona* ...('he', 'she' ...) do not only refer to people, but also to animals and things. Therefore **they** have a broader meaning. The special character of the 3rd person pronouns finds further expression in the fact that **they** more often than not refer to something which is already known from the context. **They** are therefore considered a special group of pronouns...'

This skewing in person appears also to be reflected in colloquial speech. Even speakers who are influenced by the literary language and tend to omit the personal pronouns of the 1st and 2nd person will frequently use a 3rd person pronoun. Consider the following examples:

(57) (ja) bydlu w Budyšinje
     'I live in Bautzen'

(58) woni bydla w Budyšinje
     'they live in Bautzen'

---

[37]   To be fair I need to mention that Šewc (1984: 117) only refers to 1st and 2nd person pronouns in saying that their use is reserved to emphasize the subject. On the other hand, his comment about the restricted use of pronouns refers to all persons, and he does not imply anywhere that the pronouns of the 3rd person are used more frequently.

-125-

(59) wón bydli w Budyšinje
'he lives in Bautzen'

In (57), about 70% of my informants[38] used overt *ja*, while 100% used *woni* in (58) and *wón* in (59). This was true in the perfect tense as well:

(60) (ja) sym   sej tošu kupił
     I   aux$_{1sg}$ refl bag bought$_{masc}$
     'I have bought myself a bag'

(61) wón je   sej tošu kupił
     he aux$_{3sg}$ refl bag bought$_{masc}$
     'he has bought himself a bag'

Even speakers who omit the personal pronoun in (60) use *wón* in (61). In fact, most native speakers felt that the personal pronoun of the 3rd person in sentence-initial position was obligatory. When shown literary examples with 3rd person null subjects in initial position they felt that these were "unusual" and "misleading".[39]

---

[38]   I interviewed approximately 20 native speakers of various ages and educational backgrounds.

[39]   Virtually all informants also used overt 3rd person pronouns in other positions in the sentence, as shown in (i) and (ii). However, most of them also seemed somewhat more willing to accept constructed examples without the pronoun.

(i)   je   sej wón tu tošu kupił
         aux$_{3sg}$ refl he this bag bought
         'did he buy the bag yesterday'

(i') ? je sej tu tošu kupił

(ii)   wčera je sej wón tu tošu kupił
         'yesterday he bought the bag'

(ii')? wčera je sej tu tošu kupił

The question arises whether there might be a connection between this difference in the frequency of overt pronouns and the verbal morphology of Sorbian. It is true that in the 3rd person present tense, as well as in the imperfect and past, there is no marker on the verb which specifies the gender so that the use of a personal pronoun might be blamed on the otherwise threatening ambiguity. In other words, *kupuje* ('buys') can refer both to *wón* and *wona*, just as the plural *kupuja* can refer to *wone* or *woni*. However, if this were the reason, then we could expect frequent omission of the subject pronoun in the 3rd person singular in the perfect tense, where the gender is unambiguously expressed on the *l*-participle. This expectation is not borne out, as the comparison of examples (59) and (60) shows. Moreover, in other Slavic languages which are *pro*-drop languages, such as Czech and Polish, the same situation obtains as far as the non-expression of gender in the present tense of the 3rd person is concerned. Nevertheless, in those languages null 3rd person pronominals are totally normal in unmarked contexts. In this connection I argued in 3.2. that only AGR_{pers}, and not AGR_{gend}, seems to be relevant for the identification of null subjects.

4.6.4. The competition model and the loss of *pro*-drop

Most versions of the licensing/identification framework follow the traditional wisdom that there is a direct link between the inflectional richness of a language and the availability of

null pronouns. While it seems plausible that a language would
lose the ability to have null subjects if it loses the morpho-
logical marking of agreement features on the verb, there are
languages, such as Icelandic, German, French, Rhaeto-Romance
of Switzerland and some northern Italian dialects (cf. Orešnik
(1986: 266)) which resist such a purely functional view of the
loss of null pronouns.[40] Orešnik points out that these lan-
guages developed obligatory subject pronouns when they pos-
sessed (and some still possess) full and unambiguous morpho-
logical markings on the verb.

Sprouse and Vance (1995) provide an interesting alterna-
tive to attempts made strictly within the principles and pa-
rameters framework to account for the (gradual) loss of null
pronouns in languages which were once *pro*-drop languages.
Based on data from the history of French and the Surselvan
dialect of Rhaeto-Romance they argue against a purely func-
tional view of the loss of *pro*-drop in terms of a change (de-
cline) of morphological marking on the verb. Instead, follow-
ing work by Kroch (1989), they show how a competition model of

---

[40]   Orešnik (1986) points out that in Old High German
there was a person skewing with respect to the use of null-
subject pronouns. The pronominal subjects were more frequent
in the first and second persons than in the third. He asso-
ciates the origin of unemphatic pronoun subjects with a
"strengthening", which started first in the so-called
"strong" positions. The third person, as the least marked
person is also the least likely to become strengthened.
Orešnik thus proposes that the same state of affairs with
respect to the differences among the verbal persons should
obtain in all languages. It seems that Sorbian is a counter-
example to this hypothesis, because we find a clearly higher
frequency of 3rd person pronouns in that language.

language change can account for a syntactic change over time,
one example of this being the gradual loss of *pro*-drop proper-
ties in a language. I will not present the details of Sprouse
and Vance's analysis here, but the following is a short summa-
ry of their hypothesis. Two forms of a language are said to
be in "competition" if they make exactly the same contribution
to the semantic meaning. Once a language has two competing
forms, speakers attempt to match the relative frequency of
these forms. Over time the relative rates for two competing
forms will shift in favor of the form which is parsed more
reliably, whereby "parsing" simply means grammatically based
understanding of an utterance. Once this process is set in
motion it will finally lead to the predominance of one form
and the loss of the other.

Applying this idea to the issue of null subjects, it can
be assumed that sentences with overt pronouns would be parsed
more successfully than sentences with null pronouns. However,
in order for a drift towards a high frequency of overt pro-
nouns to take place, null and overt pronouns in that particu-
lar language must have the same referential properties. Oth-
erwise, they will not be in competition and the *pro*-drop char-
acter of the language will remain stable. Consider the fol-
lowing Czech sentences:

(62) a. ja mluvím česky   'I speak Czech'

     b. e mluvím česky.   'I speak Czech'

The overt pronoun and the zero form in the two sentences above do not represent competing forms. *Ja* in Czech is a tonic pronoun and is used emphatically. The zero pronoun, on the other hand, is an atonic pronoun. The different referential properties of overt and zero pronouns in a null-subject language such as Czech can also be illustrated in the following type of construction:[41]

> (63) a. Petr$_i$ myslí, že *on$_i$/pro$_i$ má mnoho peněz
>
>   'Peter thinks that he has a lot of money'

As we have already seen in the course of the application of the null-subject criteria to Czech, the overt pronoun cannot be bound by the higher subject in the matrix clause, even though this is possible for the zero pronoun. In other words, while *pro* seems to be subject to condition B, the overt pronoun in a null-subject language seems to obey binding condition C (which boils down to the same relevant difference in the referential properties between overt and null pronouns previously pointed out by Jaeggli (1986) and Montalbelli (1984)).

In Sorbian, on the other hand, the situation is actually quite different. We demonstrated above that overt pronouns are not restricted to emphatic contexts and that embedded

---

[41]   Sprouse and Vance propose the same type of construction as a reliable criterion to identify competing forms.

overt pronouns can be coreferential with a matrix subject. Example (21f), repeated here as (64), illustrates this.

(64) Jana$_i$ je prajiła, zo chce wona$_i$/e$_i$ rady kupać hić
'Jana$_i$ said that she$_i$ would like to go swimming'

Thus there is a true competition in the relevant sense between overt and zero pronouns. If we define a null-subject language in terms of the three diagnostics in (20), a language ceases to be *pro*-drop as soon as there are atonic overt pronouns. The advantage of the competition model is that it allows for a *gradual* change in the frequency of null subjects over time, including a period of optional *pro*-drop, without being pressed to look for a change in a parameter setting. The competition model is not concerned with the formulation of precise licensing and identification conditions, though it does not explicitly deny their existence. For this sort of approach it is totally expected that there are languages in transition, during which there are both zero and overt subjects. For a licensing/identification approach there is no in-between: either null subjects are licensed or they are not. Sprouse and Vance point out that their theory is not incompatible with a parameter-setting approach such as the one presented by Roberts (1993). Their competition model is best equipped to account for a **gradual** loss of *pro*-drop properties without concomitant change in the explicitness of verbal morphology. By itself though, this model fails to explain any

sort of **sudden** change, such as the disappearance of post-verbal *pro* in French during the 17th century discussed in section 4.5.1.

Applied to Sorbian, the competition model of language change might give the following account for the loss of canonical null-subject properties. At an early stage atonic pronouns entered the Upper Sorbian language. The source for this is not clear, but German influence could certainly have played a major role as the region became bilingual.[42]

Another possible origin of atonic personal pronouns might be found in the relative order of the auxiliary and the participle in the perfect and past perfect. In contrast to Czech or Polish, in Sorbian the auxiliary has to precede the *l*-participle in neutral style. In Czech, the order *viděl jsem* ('seen have$_{1.sg}$') is normal and in fact frequently found in the initial position of a sentence. The Czech auxiliaries are second-position clitics, and the form *jsem* in *viděl jsem* is clitized to the preceding participle. Similarly, Sorbian auxiliaries are enclitics.[43] However, for Sorbian the word order option "participle - auxiliary" is not available.[44] The auxiliary has to precede the participle, so that *ja sym widźał*,

---

[42]   Sprouse and Vance do not identify the origin of atonic pronouns in Romance languages, but they doubt that a neighboring language had any role in that process.

[43]   According to Michałk (1965), Sorbian clitics prefer the "Wackernagel" position, although post-verbal position can exist if the verb occupies the second position in the clause.

[44]   This order is possible if the auxiliary is negated (cf. fn. 33).

with the overt personal pronoun preceding the clitic, is the
equivalent to the previously mentioned Czech example *viděl
jsem*. A Czech speaker would not use the same order of ele-
ments as Sorbian, unless the subject is supposed to be espe-
cially emphasized.

As discussed in section 4.4, it can be assumed that by
the 18th century overt pronouns had established their predomi-
nance over the null variant, both in spoken and written Sor-
bian. However, in the second half of the 19th century the
frequency of null pronouns suddenly increased again, mainly in
the written language, but presumably also in spoken Sorbian.
Null pronouns were artificially (and forcefully) favored by
prescriptive grammars trying to root out the German influence
in the Sorbian language and recreate Sorbian as a real null-
subject language based on the Czech model. Such a shift in
the literary language may have made the *pro*-drop phenomenon
socially significant by introducing what we might call a
"prestige factor" which artificially influenced the relative
frequency of overt and null pronouns.[45] The better educated
and well-read members of the population began to try to imi-
tate the literary language by reducing the frequency of overt
pronouns in their speech. This tradition is still alive to-

---

[45]   Sprouse and Vance mention that they abstract away
from prestige factors in the history of French null subjects.
In the 17th century French grammarians discouraged the omis-
sion of pronouns, which might have caused an artificial shift
in the relative frequency of overt and null pronouns in favor
of the latter.

day, as shown earlier in this chapter. Uneducated, simple peasant speech in literary work is imitated in part by the obvious use of overt pronouns. Moreover, all Sorbian children are taught in school to avoid the (over)use of overt pronouns. This all may explain why Sorbian today still has a competition between overt and null pronouns. Moreover, it can also account for the direction of change (overt pronouns → null pronouns), which apparently goes in the opposite direction from what the competition model predicts and is particularly striking in literary Sorbian, where null subjects became dominant. Overt pronouns do predominate in colloquial speech; however, the influence and frequency of the zero variant is great enough to keep the competition going.

Under such an account Sorbian should be analyzed as a truly *optional* null-subject language. In most environments null and overt pronouns seem to occur basically in free variation, with preference given to the overt variant. We might even try to reconcile a competition model approach to null subjects with a licensing/identification approach. Sprouse and Vance's model is meant to account for the change of *pro*-drop properties over time. They do not deal with the problem of what allows or prevents null subjects originally, i.e. they are not concerned with licensing and identification conditions. If we assume that there is some syntactic mechanism in the grammar of a native speaker which ensures the recoverability of null subjects, the question arises of whether or not

this mechanism remains intact in the process of a gradual loss of *pro*-drop. Put differently, we can ask the question of whether or not Sorbian is really a non-*pro*-drop language, in the sense that identification of referential null subjects is blocked. Perhaps the licensing/identification mechanisms have not been lost in Sorbian, so that there has not been a parametric shift as such. It is thus possible that the occurrence of null subjects in Sorbian (both literary and colloquial) is indeed a case of genuine *pro*-drop and thus does not have to be analyzed as discourse licensed ellipsis.

This approach would eliminate the problems raised at the end of section 4.6.2., specifically the question of how to account for the identification of null subjects in literary Sorbian. It also explains why purism could be so successful in (re)introducing null subjects into the literary language: the syntactic mechanism of identifying null subjects was (and is) intact in the grammar of native speakers. The null-subject diagnostics, which are based on the clear division of labor between null and overt pronouns found in canonical null-subject languages, such as Italian and Spanish, fail due to the atonic character of Sorbian personal pronouns. Interestingly, this lack of a division of labor is true not only for the colloquial language, but also for the null-subject literary language, as shown in example (56) above.

4.7. Lower Sorbian

The main focus of this chapter has been Upper Sorbian. Al-
though Lower Sorbian lays claim to a separate literary lan-
guage, it is clear that the two languages are very closely
related. It is not my intention to support one side or the
other in the ongoing debate about whether there are really two
separate Sorbian languages, but it is still worth looking at
Lower Sorbian in the context of null-subject phenomena.

Lower Sorbian developed the same ambiguity between the
2nd and 3rd person dual as Upper Sorbian. In the aorist and
imperfect paradigms this leads to a double syncretism, since
it also preserved the syncretism between the 2nd and 3rd per-
son singular from Common Slavic. Compare the Lower Sorbian
verbal paradigm below with that of Upper Sorbian in Table 1.

Table 5: Verbal Paradigm of Lower Sorbian
źělaś 'to work'

		Present/Future	Imperfect	Aorist
Sg.	1.	źělam	źělach	naźělach
	2.	źělaš	źělašo	naźěla
	3.	źěla	źělašo	naźěla
Dual	1.	źělamej	źělachmej	naźělachmy
	2.	źělatej	źělaštej	naźělaštej
	3.	źělatej	źělaštej	naźělaštej
Pl	1.	źělamy	źělachmy	naźělachmy
	2.	źělašo	źělašćo	naźělašćo
	3.	źělaju	źělachu	naźělachu

According to standard Lower Sorbian grammars, personal
pronouns in subject position are not obligatory and usually

omitted. Starosta (1991:40) gives a description reminiscent of a null subject language. According to him, personal pronouns are only used in emphatic contexts or in order to avoid misunderstandings. Janaš (1984: 171) is more noncommittal, describing their use as optional. In general, the use of overt pronouns in Lower Sorbian seems to be comparable to Upper Sorbian. The following examples are taken from Nowak's (1988) *Gutes Niedersorbisch*, which is a guide to "good" usage of the Lower Sorbian language.

(65) **my** stawamy akle ze słyńcom
    we get-up only with sun
    'we only get up with the sun'

(66) ze switanim jo **won** se wuporał
    with sunrise aux_{3sg} he refl left
    'with sunrise he set off'

(67) ga sy **ty** cora wusnuł
    when aux_{2sg} you yesterday fell-asleep
    'when did you fall asleep yesterday'

(68) **ja** som poł nocy wocuśeta lažała
    I aux_{1sg} half night awake lay
    'I lay awake half the night'

(69) **my** wjele nowego njezgonijomy
    we much new not-hear
    'we don't hear much new'

Note that in all these examples there would not have been any risk of ambiguity had the pronoun not been used. Neither are these sentences examples for overt pronoun use in emphatic or

contractive contexts.[46] Thus, Lower Sorbian is not a canonical null-subject language in the sense defined in previous chapters. The application of the other two null-subject criteria leads to the same conclusion. Overt Lower Sorbian *woni* ('they') can have arbitrary reading, thus contradicting diagnostic (20c). Similarly, an overt pronoun may have a bound variable reading, causing diagnostic (20b) to fail. This is illustrated in the examples below.

---

[46] Word order in Lower Sorbian has not been studied in as much detail as in Upper Sorbian. In principle the word order pattern seems to conform more or less with that of Upper Sorbian. Waurick (1968: 126) writes:

> Wie im Obersorbischen herrscht auch im Nieder-sorbischen der Grundsatz, daß das nichtzusammengesetzte verbale Prädikat (verbum finitum) in der Regel am Satzende steht. Jedoch scheint diese Regel etwas großzügiger angewendet zu werden als im Obersorbischen, denn verhältnismäßig häufig tritt Zweit- oder Mittelstellung des nichtzusammengesetzten verbalen Prädikats auf. Ähnlich wie im Obersorbischen stehen bei einem zusammengesetzten verbalen Prädikat die finite Form des Hilfverbs gewöhnlich in Zweit- oder Mittelstellung und das *l*-Partizip bzw. der Infinitiv am Ende des Satzes. Jedoch gibt es auch hier wahrscheinlich zahlreichere Ausnahmen als im Obersorbischen.

> 'As in Upper Sorbian, in Lower Sorbian there is a principle according to which the non-compound verbal predicate (finite verb) usually is in final position. However, this rule seems to be applied less strictly than in Upper Sorbian, because relatively often the non-compound verbal predicate occurs in second or middle position. As in Upper Sorbian, the finite form of the auxiliary in compound verbal predicates is usually in second or middle position, and the *l*-participle or infinitive at the end of the sentence. However, there are probably again more exceptions than in Upper Sorbian.'

-138-

(70) a. w Londonje **woni** zasej stajaškuju
in London　they again strike

'in London they are again on strike'

b. we Łužycy **woni** maja rědne stare nałogi
in Lusatia they have nice　old　　customs

'in Lusatia they have nice old customs'

(71) a. kuždy nan　　se　　mysli, až **won** ma mudrego syna
every father refl. thinks that he has smart son

'every father thinks that he has smart son'

b. kuždy płosaŕ se　　nażeja, až　　se　**won** skoro
every beggar refl hope　　　that refl he　soon

rozbogaśijo
becomes-rich

'every beggar hopes that he will soon become rich'

c. Frido wě,　　až　**won** njebužo　ains krydnuś
Frido knows that he　not-will A　　receive

'Frido knows that he won't get an A'

As above with Upper Sorbian, it should be emphasized that
there are also many instances of null-subject pronouns in
Lower Sorbian, so that the overt and null variants seem to be
in free variation.　Below are some examples with the null
variant, taken from Nowak's (1988) book.

(72) som　jogo poznł　　　　wot　jog nejlěpšego boka
aux$_{1sg}$ him　got-to-know from his best　　　　side

'I met him from his best side'

(73) cas　žywjenja njejsom　　jogo poznał
time life$_{gen}$　　not-aux$_{1sg}$ him　met

'I have never met him in my entire life'

(74) na dobru gluku smy　　do kjarcmy zajšli
at good　luck　aux$_{1pl}$ to pub　　went

'we went to the pub on the off-chance'

The use of overt pronouns is widespread in the earliest
texts, such as the Lower Sorbian Bible translation from 1548
by Miklawuš Jakubica. The translation of the New Testament by
Jan Bogumił Fabricius in 1709 shows a very frequent use of
overt personal pronouns. A direct comparison of one passage
(Luke 10:31-35) reveals six overt pronouns in the Jakubica
translation and nine overt pronouns, (including one overt
expletive), in the Fabricius translation. The Fabricius
translation is cited below.

(75) Luke 10:31-35; Fabricius 1709

31. **Wono** pak pšifoźecy se stanu, až jaden mjeršnikna
    tej samej droze doloj źešo, a gaž jogo wupyta, źešo
    won mimo.
32. A teje rownosći tež lewita, gaž **won** na to město
    pšiźe, a jogo wiźešo, źešo **won** mimo.
33. Jaden Samariski pak ducy na tej droze, pšiźe tam, a
    gaž **won** jog wiźešo, bu jomu jogo luto.
34. A pšistupjecy k njomu hobalowašo **won** jogo rany, a
    lejašo nutš wolej a wino a sajźi jogo na swojo
    zwěrje, wjěźešo jogo do gospody a zažarbowašo jogo.
35. A na drugi źeń dalej ducy wujmje **won**, a da jej tomu
    gospodarjoju, a žašo k njomu: "Zažarbuj jogo, a
    buźoš li njeco wjecej naloźyš, cu **ja** tebe jo
    zaplaśiš, gaž **ja** pšidu".

Similarly, in Matthew 2:2-18, i.e. the same passage we used
before for comparison between the different versions of Upper
Sorbian Bibles,[47] there are nine overt, non-emphatic pronomi-
nal subjects in the Jakubica and ten in the Fabricius version.

The use of overt pronouns is also evident in the every-
day language of the 19th century, as illustrated in the fol-

---

[47]   See section 4.5.

lowing excerpts taken from Nepila's "Głodne lěto". Nepila had very little formal training in writing and his autobiographical stories are said to closely reflect his native dialect of Lower Sorbian. Because of this, he is considered the first *ludowy spisowaćel* ('writer of the people').[48]

> ... ja cu Wam tež jenu nowu powjesć prjodk ćinjesć wót togo drogego lěta tedym hako ja som mały był a kajku ja jědź jědł som tedym we drogem a głodnem lěće ...

> 'I want to give you a new account of that year, back then when I was little, and what kind of food I ate during that year of increased prices and starvation ...'

> ... da **wóni** su tež wopytowali wót šogo jěsč, ale **wóni** su jare mało zjědli; **wóni** su jěsć wopytowali, ale **wóni** njejsu mógli teje jědźi jěsć. **Wóni** wopytowali tež derje su, ale **wóni** su šycko roztylja rozchytali ...

> 'and **they** too tried to eat from everything, but **they** ate very little; **they** tried to eat, but **they** couldn't eat that food. **They** had a taste of it, but **they** threw everything around ...'

The *serbske narodne wozrodženje* did not reach Lower Lusatia until the end of the 19th century. Even then, the striving for purism in the literary language was not quite as strong as in Upper Sorbian. The version of the Lower Sorbian New Testament published in 1895 in Halle was still Fabricius's (1709) translation. In Upper Sorbian, on the other hand, a new Bible translation by Lusčanski and Hórnik was published which was "cleansed" of virtually all subject pronouns which

---

[48]   Examples cited from Lorenz (1981: 94-98).

were so abundant in Swětlik (1690).[49] The Upper Sorbian exam-
ple below is the same passage as the Lower Sorbian in example
(75). There is only one overt personal pronoun in (76), com-
pared to nine in (75).[50]

(76) Luke 10:31-5; Lusčanski and Hórnik 1896

31. Sta pak so, zo měšnik po tom samym pućudźěše, a
    wuhladawši jeho, přeńdźe nimo.
32. Tak tež lewita, jako tam porno přińdźe a jeho
    wuhlada, přeńdźe nimo.
33. Jedyn Samaritan pak tam pućujo přińdźe porno njemu,
    a wuhladawši jeho bu z miłosću hnuty.
34. A přistupiwši, zawjaza jeho rany, wolij a wino na
    nje linywši, a sadźi jeho na swoje skoćo, dowjedźe
    jeho do hospody, a zastra jeho.
35. A nazajtra wućahnywši dwaj dźesatnikaj da jej
    hospodarjej a praji: Staraj so za njeho, a štožkuli
    wjacy wuložiš, ja, hdyž so wróću, zapłaću tebi.

In summation, it appears that Lower Sorbian behaves more or
less like Upper Sorbian with respect to the use of overt and
null-subject pronouns.

4.8. Summary

In this chapter I discussed the status of Upper Sorbian with
respect to null-subject properties. Most grammarians describe
Upper Sorbian as a canonical null-subject language, and the
literary language in general avoids overt subject pronouns

---

[49]   Also compare section 4.4.

[50]   The personal pronoun in Luke 10:35 of the Lusčanski
and Hórnik version is also present in the Latin Vulgate:

    Et altera dir protulit duos denarios, et debit
stabulario, et ait: Curam illius habe: et quodcumque
supererogavis, ego cum rediero reddam tibi.

(with the possible exception of 3rd person pronouns). However, the situation is quite different in the spoken language. There, nonemphatic overt pronouns occur very frequently. I argued that the gap between literary and colloquial languages is the result of a strong movement for language purism which started in the second half of the 19th century and strove to eliminate elements of German origin from the Sorbian literary language. An application of the three null-subject diagnostics to Upper Sorbian revealed that Upper Sorbian does not show the division of labor typical for a real null-subject language. I introduced three potential accounts of why Sorbian might be different from other West Slavic languages with respect to its null-subject properties.

Two of the accounts were formulated within the licensing and identification framework. One possible solution to the question of what might be blocking the identification of referential pronouns was explained in terms of Sorbian crossing the threshold of allowable syncretisms required for functional richness of a verbal paradigm. The other possible account for the blockage of identification was derived from the peculiar SOV word order of Sorbian. Sorbian is underlyingly (and in most cases also at S-structure) V-final. I speculated that the problem for identification may have something to do with the non-adjacency of AGR and [Spec,AGRP].

The third proposal was based on an application of the competition model of language change to the issue of the grad-

ual loss of *pro*-drop, as proposed by Sprouse and Vance (1995). According to this analysis, a null-subject language will start moving towards the complete loss of *pro*-drop as soon as atonic overt pronouns enter into competition with null pronouns. Atonic pronouns in Sorbian may have come about under German influence and/or as a result of the relative order of the participle and its auxiliary in the perfect and past perfect tenses.

All three accounts were able to explain why Sorbian is not a null-subject language in the strict sense (despite the fact that they had difficulties in accounting for certain data and/or made counterintuitive predictions). I argued that any strict application of the licensing/identification framework encounters problems in accounting for the optionality of Sorbian null/overt subjects. Thus, only the competition model analysis seemed flexible enough to make allowances for the specific character of Sorbian as an **optional** null-subject language, unless we want to assume that the presence of refer-ential, atonic pronouns and the availability of an identifica-tion mechanism are not mutually incompatible.

5. THE STATUS OF OVERT EXPLETIVES IN CZECH AND SORBIAN

5.1. Licensing/identification and the typology of languages

In the previous chapters we were mainly concerned with the occurrence or lack of overt referential subject pronouns. However, the lack of overt referential pronouns in a language does not necessarily mean that the language allows no null subjects at all. In the review of the development of the *pro-drop* parameter in Chapter 2 we saw that the existence of mixed languages (which have null non-referential (expletive) subjects but require overt thematic subjects) has suggested that the null-subject parameter cannot be monolithic as originally assumed.[1] As we have seen in Chapter 2, one standard approach to the problem of mixed languages is to capitalize on the distinction between licensing and identification conditions on null subjects. The licensing conditions specify the environment in which null subjects are allowed to occur. All null subjects, independent of their thematic status, must be formally licensed. However, **only** referential null subjects re-

---

[1] Within government binding theory, expletive subjects are noun phrases in subject position which are not assigned a thematic role (such as agent, patient etc.). They are thus different from referential pronouns on the semantic level. There are also a number of syntactic differences. For example, in contrast to referential pronouns expletives cannot be questioned (except in tag questions) and cannot be coordinated with another NP. This is illustrated in the examples below:

(i) It is cold. What is cold? (it = the beer)
     It is raining. *What is raining?

(ii) It and the coffee are cold. (it = the meal)
      *It and the coffee are cold. (with it = expletive)

quire **additional** identification to ensure unambiguous recov-
erability of their pronominal content. Null expletives, on
the other hand, merely need to be licensed. In a mixed lan-
guage *pro* is formally licensed to occur in a certain position
but is not identified. It follows that in such a language *pro*
is restricted to expletives, which, due to their non-referen-
tial status, do not need to be identified. This type of sys-
tem predicts the three different language types in Table 1.

Table 1: Possible language types (I)
   i.   canonical *pro*-drop languages
   ii. mixed type: overt referential pronouns; null
        expletives
   iii. canonical non-*pro*-drop languages

The essence of the licensing/identification dichotomy is
that it is much easier for expletives than for referential
pronouns to be null, since their pronominal content does not
need to be identified. Identification is more "difficult"
than licensing and can perhaps be understood as "more
complete" licensing.[2]

The licensing/identification framework thus predicts that
a language type with null referential and obligatorily overt
expletive pronouns cannot exist. Hyams (1986: 71-74) points

---

   [2]   See also 2.3. about the practical impossibility of
identification without licensing.

out that this prediction follows from Chomsky's avoid pronoun

principle (1981: 65):[3]

> (1) Avoid Pronoun Principle
>
> Avoid lexical pronominal if a null pronominal is
> possible.

A language which allows null-referential subjects must be able

to license null subjects in general. If *pro* is licensed, then

a language will have null expletives. Indeed, Hyams extended

this prediction to state that not only will expletives in such

a language be allowed to be zero, they **must** be zero. In other

words, languages which license phonologically null subjects

will not contain lexical expletives at all. Note that even in

a canonical null-subject language the subject pronoun is some-

times lexicalized. However, its use is reserved for special

functional purposes, such as emphasis, contrast or change of

discourse topic, and also avoidance of ambiguity, and thus is

not a violation of the avoid pronoun principle. In certain

contexts the overt variant cannot be "avoided", because it

carries some additional meaning which the zero variant is

lacking. The use of overt subject pronouns in a canonical

null-subject language is thus determined by **functional** consid-

erations. Expletive pronouns, on the other hand, are by defi-

nition void of semantic content, and it is a standard assump-

---

[3]   The status of a language as an **optional** *pro*-drop
language, as was suggested for Sorbian in section 4.5.4., may
jeopardize the general validity of Chomsky's principle.

tion that they cannot be used for functional purposes. Thus, there should never be a reason to use an overt expletive subject in a null-subject language. Based on this idea, language acquisition theorists such as Hyams (1987: 74) have argued that the existence of overt expletives in a language serves as an unambiguous signal to the language learner that the language does not allow empty subjects of any kind. According to Hyams, the presence of overt expletives in a language is a "trigger" which induces the restructuring of the original parameter setting [+*pro*-drop] which is supposed to be the unmarked value of the null-subject parameter of UG.

Franks (1990) provides an interesting alternative approach to null subjects which places emphasis on different case assignment properties. He proposes two (more or less independent) parameters:

(2) Nominative Case features must be realized.

(3) Nominative Case assignment cannot take place at LF.

Descriptively, Franks calls (2) the overt expletive subject parameter [OESP], while (3) is the overt subject parameter [OSP]. In contrast to the previous system, these parameters do predict **four** different language types:

Table 2: Possible Language Types (II)

    i.   [-OSP,-OESP]

         optionally overt thematic subjects,
         obligatorily zero expletives

    ii.   [+OSP,-OESP]

         obligatorily overt thematic subjects,
         obligatorily zero expletives

    iii.   [+OSP,+OESP]

         obligatorily overt thematic subjects,
         obligatorily overt expletives

    iv.   [-OSP,+OESP]

         optionally overt thematic subjects,
         optionally overt expletives

This notation is somewhat misleading, since [+OSP] means obligatorily overt thematic subject, but [+OESP] can mean obligatorily overt or optionally overt expletive subjects, depending on the setting for the [OSP] parameter. Similarly, [-OSP] means optionally overt thematic subjects, while [-OESP] means obligatorily zero expletives. The combination [+OESP] [-OSP] is thus not an instantiation of the fourth "forbidden" type[4] under the licensing/identification approach. Instead, it means that while nominative case must be discharged to subject position there is the option to assign nominative at LF rather than at S-structure. These asymmetries between [OESP] and [OSP] are caused by the interaction of both nominative case parameters in (2) and (3) with the so-called case filter, according to which an NP must have case in order to be

---

[4]   Forbidden type: null referential subjects, obligatorily overt expletives.

visible. Specifically, as mentioned by Franks (1990), there are two sides to this requirement: all overt NPs must be visible at PF; only theta-marked NPs must be visible at LF.

The main focus of this chapter is to discuss how Czech and Sorbian data are compatible or incompatible with the licensing/identification framework. Therefore, I will not follow this interesting line of argument any further here and will come back to Franks's (1990) proposal only as an aside. Below I will show how overt non-referential subject pronouns in spoken Czech, although in contradiction with Hyams's prediction about the total lack of expletives in any null-subject language, can be predicted by the licensing/ identification system without a violation of Chomsky's avoid pronoun principle. This is followed by a discussion of Upper Sorbian expletive-like pronouns. Because of the distributional restrictions on the occurrence of such pronouns I will argue that they do not cause any problems for the licensing/identification approach either.

5.2. Spoken Czech[5]

The standard, written language (*spisovná čeština*), was codified at the beginning of the 19th century by Josef Dobrovský. Spisovná čeština is based on 16th-century Czech, under the assumption that the German/Habsburg repression after 1618 had corrupted "pure" Czech. This romantic view of historic Czech

---

[5]   The hypotheses and data in section 5.2. and 5.3. have, in part, been previously presented in Lindseth and Franks (1994) and Lindseth (1994).

ignored several changes which had been developing in the Czech language since the 16th century. The written language was in many respects already archaic at the point in time when it was recreated. It is, as the name *spisovná* indicates, a written and not a spoken language. The spoken language developed in a natural way without the artificial intervention of philologists. Thus there is a historic and linguistic gap between literary (written) Czech and spoken Czech. It is important to understand that the difference between the two cannot be compared to the difference between the written and spoken variants of other languages. As Townsend (1990) points out, all languages have layers of slang, dialect words, jargon. In Czech too, these layers are incorporated in the spoken language, and various emotive and subjective factors are relevant. However, there are large numbers of **neutral** pronunciations, morphological forms, syntactic constructions and words used by educated speakers in everyday conversation which differ from those prescribed by the literary language.

The term "spoken Czech" is somewhat problematical. It is sometimes used to refer to Common Czech (*obecná čeština*), which is an interdialect spoken in Bohemia and Southwestern Moravia, or it can refer to Commonly Spoken Czech (*běžně mluvená čeština*), sometimes also referred to as Colloquial Czech. Several linguists (Hammer 1985, Short 1993, Townsend 1990) see the literary language (*spisovná čeština*) and Common Czech (*obecná čeština*) as the two basic codes, while Commonly Spoken

Czech (*běžně mluvená čeština*) is somewhat of a transitional stratum. Short describes it as a "channel" by which "upwardly mobile features" of Common Czech can enter the standard language. Hammer (1985) describes it as a blend of literary and Common Czech, whereby Common Czech is the foundation, and speakers switch to the literary forms for functional purposes of emphasis, irony, intimacy, etc.

The following discussion is not focused on the relationship between the various forms of the Czech language. Instead, I will investigate only one syntactic phenomenon which can be found in the spoken language (in both *obecná čeština* and *běžně mluvená čeština*).

5.2.1. Syntactic status of overt expletives

The Czech personal pronouns are *já* ('I'), *ty* ('you'), *on, ona, ono* ('he, she, it') in the singular, and *my* ('we'), *vy* ('you'), and *oni* ('they') in the plural. In stylistically unmarked discourse, Czech (both literary and spoken) omits unstressed pronominal subjects. This is illustrated in the passage below taken from Heim (1982: 151).

> (4) 1. Potkal jsem ho, právě když vycházel z
>   knihkupectví na Karlově náměstí.
>   2. Chodili jsme spolu skoro hodinu.
>   3. Najednou mi řekl, že musí na schůzi, že
>   prý už tam - tj. na druhé straně města - na něj
>   někdo čeká, a odběhl.
>   4. "Nikdy ho asi nepochopím", řekl jsem si.
>
>   1. (I) met him, just when (he) came out of the
>   bookstore at Charl's square.
>   2. (We) walked together for almost an hour.

3. Suddenly (he) told me, that (he) must go to a
   meeting, and that someone is waiting for him
   there, that is on the other side of town, and
   (he) ran off.
4. "(I)'ll never understand him",(I) said to myself.

It has been shown in Chapter 3 that, with respect to the
other two null-subject criteria, Czech also qualifies as a
canonical null-subject language. Recall that, in Czech, overt
pronominal subjects cannot act as bound variables, as illus-
trated in (5) and (6).   Nor can they be interpreted as arbi-
trary in reference, as shown in (7).

(5) každý učitel, si   myslí, že on.$_{i/j}$ má hloupého syna
    every teacher refl thinks that he has stupid son
    'every teacher thinks that he (=some specific
    person) has a dumb son'

(6) Petr$_i$ ještě stále   miluje tu ženu, kterou si
    Peter still continue loves the women whom   refl

    on.$_{i/j}$ před pěti lety vzal.
    he   ago 5 years took

    'Peter still loves the woman whom he (=that other
    man) married 5 years ago'

(7) v Německu *oni rádi   pijou české pivo
    in Germany they gladly drink Czech beer
    'in Germany people like to drink Czech beer'

In section 4.3. following Faska (1964), a list of three
environments was given in which Sorbian strongly prefers overt
thematic subjects. They are repeated below under (8).

-153-

   (8) An overt pronominal subject is preferred in Sorbian
        a) if direct or indirect speech is being introduced
        b) if the matrix subject is a demonstrative
        c) if the matrix subject is a full NP

In Czech, on the other hand, overt pronouns are not at all privileged in those environments, but instead are avoided as shown in examples (9-12). Environment (8c) directly violates one of the null-subject criteria. Example (9) below is thus related to the sentence type illustrated by (5) and (6) above. Sentences (10) and (11) contrast with (8a), and (12) contrasts with (8b).

   (9)   mí rodiče řekli, že přijdou   zítra
        my parents said that will-come tomorrow
        'my parents$_i$ said, that (they)$_{i/j}$ will come tomorrow'

   (10) řikal, že   je   unavený
        said$_{masc}$ that aux$_{3sg}$ tired$_{masc}$
        '(he) said that (he) is tired'

   (11) "nebojím   se   ničeho", řekl
        not-afraid$_{1sg}$ refl nothing   said$_{masc}$
        '"(I) don't fear anything", (he) said'

   (12) ta   řekla, že   je   unavená
        that$_{fem}$ said$_{fem}$ that aux$_{3sg}$ tired$_{fem}$
        'that one said that (she) is tired'

These facts further support the claim that Czech is a canonical null-subject language, in contrast to Sorbian.

However, elements which appear to be overt expletives do occur frequently in spoken Czech.[6] Overt expletives are not found in stylistically neutral written texts; however, they can occur in text types which in general try to imitate the spoken language. Schalich (1970: 78-79), for example, shows how expletive pronouns are used as "expressive means of language" in modern Czech comedies.

There have been very few attempts thus far to account for this phenomenon. Trávníček (1962) deals with it from a diachronic perspective, while Koenitz (1988) and Hirschová (1984) try to provide an account within the framework of functional sentence perspective. Both analyze the expletive as either belonging to the thematic part of the sentence or marking the actual (very general) theme of the sentence. Franks (1995) and Lindseth (1994) are the only analyses within government binding, and Woolhiser (1994) provides a general survey of the syntactic and semantic characteristics of expletive-like pronouns in several Slavic languages.

Expletives in spoken Czech are optionally overt, so this is not an instantiation of the fourth language type characterized as impossible in the identification/licensing model. However, the linguistic facts seem to contradict previous predictions, e.g. Hyams's, about the complete absence of overt

---

[6] Townsend (1990) points out that there is a higher frequency of overt referential pronouns in the spoken language. This is not too surprising, considering that more emotive/subjective factors enter into this type of speech.

expletive subjects in a language with null referential pronouns.
The examples below illustrate this phenomenon, which poses an
interesting problem for the licensing/identification framework.
In none of the sentences is the subject assigned a thematic role.
The subject position thus seems to be occupied by an expletive
pronoun. The form of the Czech expletive pronoun *(v)ono* is iden-
tical to the personal pronoun of the third person singular neuter.[7]

> (13) a. **vono** je chladno
>> 'it is cold'
>
> b. **vono** ještě prší
>> 'it is still raining'
>
> c. **vono** se blýská
>> 'it is lightening'
>
> (14) a. **vono** se   tam   nepracuje
>> it   refl there not-works
>>
>> 'no work is being done there'
>>
>> (cf. Es wird dort nicht gearbeitet)
>
> b. **vono** se   tu   tancuje
>> it   refl there dances
>>
>> 'dancing is going on there'
>>
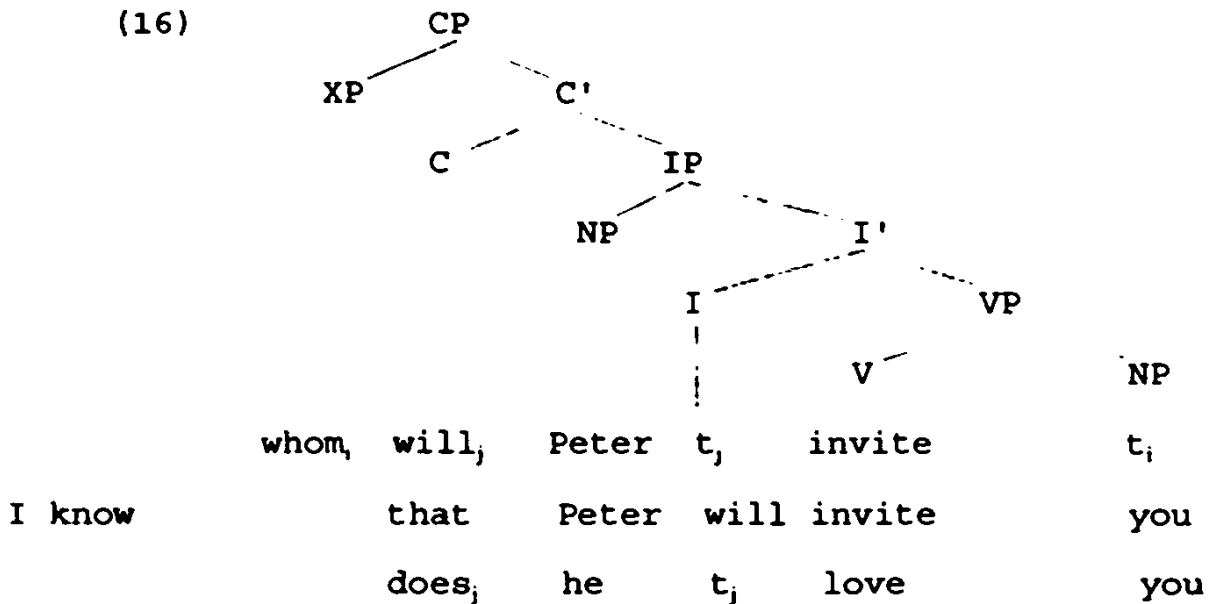>> (cf. Es wird dort getanzt)

---

[7]   One obvious phonological feature of Common Czech
in contrast to Standard Czech is the use of a prothetic con-
sonant [v] before initial *o* in most native Czech words, in-
cluding the third person pronouns. *Ono* thus becomes *vono*.
Hammer (1985) points out that in colloquial Czech this fea-
ture is used by virtually all speakers when it comes to pro-
nouns and prepositions, but that its use in other words is
more sensitive to the topic of the conversation, the type of
words, the background of the speaker, etc.   Townsend
(1990:36) calls the prothetic shift *o > vo* "the most impor-
tant SPC (Spoken Prague Czech) feature and certainly the most
difficult to analyze", presumably because of the multitude of
factors which influence its use.

(15) a. **vono** je možný,   že   tam   nepřijede
       it   is possible that there not-comes
       'it is possible that he/she doesn't get there'

   b. **vono** je přece nutný,   abysme jí   to   řekli
       it   is yet   necessary $aux_{1pl\ subj}$ her that $said_{pl}$
       'it is after all necessary that we tell her'

   c. **vono** se   zda, že   to   nebude   lehký
       it   refl seems that this not-will-be easy
       'it seems that this is not going to be easy'

An immediate question is whether or not these expletives are really subjects. Of course, the answer to this question will depend on our definition of subject. Within Government Binding theory the notion of subject is defined configurationally as the NP which occupies the specifier position of IP.[8] First of all, I assume that expletives have the categorial status of NPs and are not some sort of particle or adverbial. They are thus eligible to occupy subject position. This assumption is based on the fact that their form coincides with that of the third person neuter personal pronoun, and, in the examples above, they seem to correspond to overt non-referential subject pronouns in non-null-subject languages (*it* in the English translation of the examples above). If the expletive is in subject position, then we would expect it to be compatible with material in C or [Spec,CP], as is the case for a normal

---

    [8]   Under the split-INFL hypothesis, the subject position would be [Spec,AGRP]. However, this distinction has no bearing on the current discussion.

subject. The tree in (16) illustrates this for some English examples.[9]

(16)

	whom$_i$	will$_j$	Peter	t$_j$	invite	t$_i$
I know		that	Peter	will	invite	you
		does$_j$	he	t$_j$	love	you

For one thing, expletives in Czech seem to be more or less restricted to sentence initial position, which could be taken to indicate that they occupy the specifier of CP instead of [NP,IP]. If they are not structural subjects, then they should not cause any trouble *per se* for the licensing/identification framework, since there could still be an empty NP (*pro*) in [Spec,IP] (with the non-referential element occupying a different position "further up" in the sentence).[10]

---

[9]   In English there is a restriction that a *wh*-word in [Spec,CP] and an overt complementizer under C cannot co-occur. In the literature this constraint is referred to as the "Doubly-filled COMP filter", dating back to a time when [Spec,CP] and C were not clearly distinguished. In several other languages, e.g. in Dutch and in Bavarian German, the sequence of a *wh*-word and a complementizer is possible.

[10]   For example, one subtype of German non-referential *es* as in (i) seems to be restricted to [Spec,CP], and is thus incompatible with *wh*-words such as *wann* ('when'), as shown in

The examples below were judged by native speakers for
their grammaticality. The complementizer *protože* in (17-19),
když in (20) and the preposed verb *nepracovalo* in (21) occupy
C, while the question words in (22-24) are in [Spec,CP].

(17) nemůžu jít ven,   protože **vono** tam   prší
      not-can go outside because it   there rains
      '(I) can't go outside, because **it** is raining there'

(18) nikdo tu   nebyl, protože **vono** se   tu
      nobody there not-was because it   refl there

      včera   nepracovalo
      yesterday not-worked~nax~

      'nobody was there because there was no work being
      done yesterday'

(19) protože **vono** se zda,   že   to   nebude   lehký
      because it   refl seems that this not-will-be easy

   'because it seems that this is not going to be easy'

(20) ?? nemůžeme tady být, když **vono** se   tu   nepracuje
         not-can~ig~here be   if   it refl here not-works

      'we can't be here, if there is no work being done'

(ii). (See also the discussion in section 5.3.3.)

   (i) **es** wurde wie immer viel gelacht und gesungen
      'there was as usual a lot of singing and dancing
      going on'

   (ii) *wann wurde **es** wie immer viel gelacht und gesungen?

Under this analysis sentence (i) actually has an empty NP
(*pro*) in subject position, as illustrated in (iii).

   (iii)

es wurde~i~ *pro* gelacht   t~i~
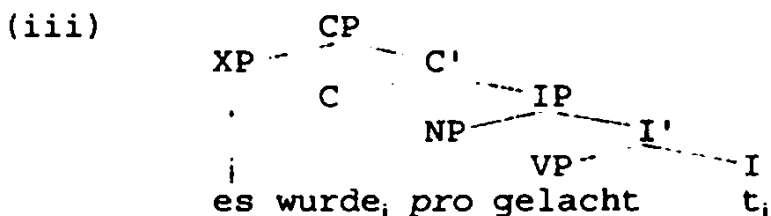
(21) ?? nepracovalo se   **vono** včera?
      not-worked_{neu} refl it   yesterday
      'wasn't there any work done yesterday'

(22) *kde je **vono** chladno
     'where is it cold'

(23) *kdy **vono** prší/prší **vono**
      when it rains/rains it
      'when is it raining'

(24) *proč **vono** se   tu   včera   nepracovalo
      why it   refl here yesterday not-worked_{neu}
      'why wasn't there any work done yesterday'

At first glance, the results of this investigation seem to contradict each other. Native speakers were quite willing to accept complementizers (especially *protože*) preceding *vono*; however, constructions with material in [Spec,CP] were rejected. As I will discuss below in section 5.2.2., the marginality or unacceptability of sentences (20-24) may be due to pragmatic rather than syntactic reasons. I thus claim, based on the compatibility between *vono* and material in the complementizer position C in examples (17-19), that expletives in Czech are indeed in subject position.[11]

The Czech expletive is not only used in impersonal constructions, such as impersonal passives and weather verbs, but

---

[11]   Within the typology proposed by Franks (1990), spoken Czech is thus an example for [-OSP,+OSP].

it can co-occur with overt referential NP subjects, as shown
in (25-32).[12]

(25) **vona   ta   myšlenka** má něco   do sebe
     expl.$_{fem}$ this thought$_{fem}$ has something to itself
     'this thought has something to it'

(26) **vona**   se   **Mařenka** stará   vo všechno
     expl.$_{fem}$refl   Mařenka$_{fem}$ takes-care of everything
     'Mařenka takes care of everything'

(27) **vona**   se   tu   naskytla **ta vosoba**
     expl.$_{fem}$ refl there appeared that person$_{fem}$
     'there appeared that person'

(28) **von**   mu   **otec** všechno   dovolí
     expl.$_{masc}$ him   father everything allows
     'father allows him everything'

(29) **von**   ho   nemine   **trest**
     expl.$_{masc}$ him   neg-spare   punishment$_{masc}$
     'punishment won't spare him'

(30) **vono**   tam   bylo   **moc lidí**
     expl.$_{neut}$ there was$_{neut}$ many people$_{gen.pl}$
     'there were many people there'

(31) **vono**   se   ti   **to lhaní** jednou vymstí
     expl.$_{neut}$refl. you$_{dat}$ this lying$_{neut}$once   revenge
     'this lying will come back to haunt you sometime'

(32) **von**   byl dřiv **vchod**   z   druhý strany
     expl.$_{masc}$was once entrance   from other side
     'there used to be an entrance from the other side'

---

[12]   Example (25) is constructed, example (26) is taken
from Koenitz (1988), example (27) from Schalich (1970), example
(28) and (29) from Siebenschein's dictionary (1983), examples
(30) and (31) from Hirschová (1984), example (32) from
Townsend (1990). All examples were tested on and accepted by
native informants.

There is no intonational break between the pronoun and the rest of the sentence, which indicates that this is not some type of left-dislocation.[13] Moreover, the expletive in such constructions agrees with the associated referential NP in pronominal features. We thus have feminine *vona* in (25), (26), and (27), masculine *von* in (28), (29), and (32), and neuter *vono* in (30) and (31). Example (30) is interesting with respect to agreement. It contains a quantified NP in the genitive case, so that default neuter singular agreement appears on the pronoun. The finite verb *bylo* is also in the neuter singular form. Both the expletive and the thematic NP appear in the nominative case.

Expletives in this construction are for the most part compatible not only with overt material in C, as in examples (33) and (34), but also in Spec-CP, as shown in (35-37)[14].

(33) myslím, že **vona** ta tvoje **myšlenka** má něco do sebe
'I think that your thought has something to it'

---

[13] In left-dislocation constructions, an NP is adjoined to the left of IP and picked up again by a coindexed NP, as shown in the example below.

binding principles$_i$, I cannot figure them$_i$ out

Coindexation of left-dislocation constructions is not a violation of principle B or C (a pronoun has to be free in its governing category; an R-expression has to be free everywhere), since the binding theory applies to A-binding only and the antecedent in the left-dislocation construction occupies an A' position.
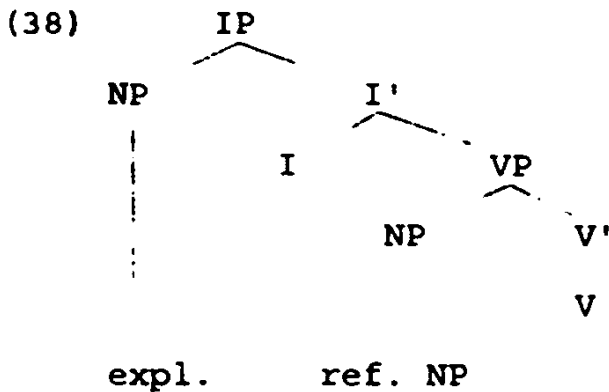
[14] For glosses compare examples (25)-(28).

(34) ... protože **vona** se **Mařenka** stará vo všechno
     'because M. takes care of everything'

(35) najednou **vona** se tu naskytla **ta vosoba**
     'suddenly that person appeared there'

(36) kdy **vona** se tu naskytla **ta vosoba**
     'when did that person appear here'

(37) proč **von** mu **otec** všechno dovolí
     'why does dad allow him everything'

Based on these facts, I conclude that the expletive is in [Spec,IP]. The unexpected compatibility of the expletive with material in [Spec,CP] in this construction type (in contrast to the *vono prší/vono se tu nepracuje* type of construction) will be discussed in 5.2.2.

The next question which arises is which structural position the referential nominative NPs in this "double subject" construction type occupy. According to Sportiche's (1988) VP-internal subject hypothesis, all arguments, including the subject, are projected in the VP. Thus there is another subject position, namely [Spec,VP], which is assumed to be the D-structure position for all thematic subjects. I propose that the referential subject NPs in the examples under (25-37) actually remain in Spec-VP at S-structure. This is illustrated in (38).[15]
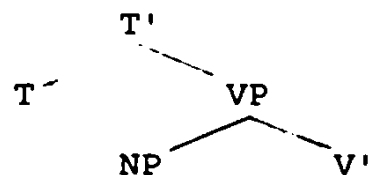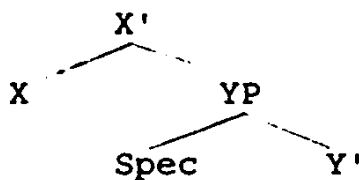
---

[15] At LF the referential NP moves up to Spec-IP to adjoin (or replace) the expletive, so that no violation of the principle of full interpretation occurs, according to which LF may only contain elements which contribute to the semantic interpretation.

(38)

```
             IP
           /    \
    NP          I'
    |         /    \
    |       I        VP
    |             /    \
    |          NP        V'
    :                     
                          V

  expl.       ref. NP
```

The expletive and the referential NP associated with it pre-sumably form some sort of CHAIN at S-structure. The members of this chain share pronominal features and case.

Alternatively, nominative case assignment to the expletive and the thematic pronouns could occur independently. As pointed out before, within recent analyses it has become standard to assume that what was formerly IP should be split into several specialized functional projections of V, such as AGRP, TP, NegP etc. Let us further assume with Roberts (1993: 27) that either AGR or T can assign nominative case as a matter of parametric choice. According to Roberts, Agr can assign case under spec-head agreement or government ECM; however, T can only assign case under government ECM.[16] It is thus conceiv-

---

[16]   ECM = Exceptional Case Marking. This case marking is "exceptional" because the case assigning head X assigns case to the specifier position of a maximal projection YP. Maximal projections normally constitute barriers for case assignment from outside.

```
        X'                              T'
      /    \                          /    \
   X        YP                     T        VP
          /    \                          /    \
       Spec     Y'                     NP        V'
```

-164-

able that in some languages, including under this analysis
Czech, T assigns nominative case under government of the sub-
ject NP in Spec-VP. At the same time, AGR is able to assign
nominative to its specifier, occupied by the expletive, under
spec-head agreement.[17] This alternative explanation of how
nominative case assignment can be accounted for in the Czech
double-subject construction is shown in (39).

(39)   ArgP

      XP   Agr'

          Agr   TP

              T   VP

                  NP   V

  expl.   ref. NP

---
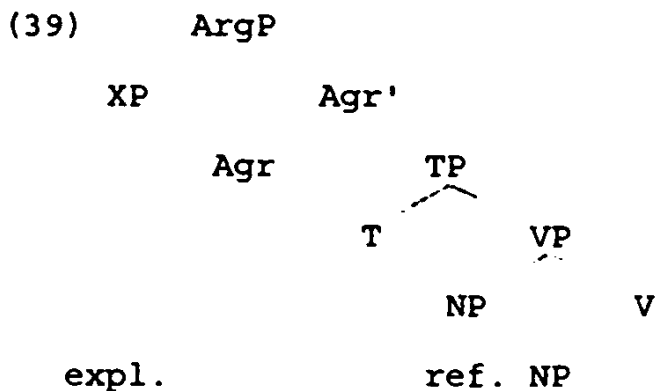
[17]   In principle, we could expect that a subject can
use different Spec positions as landing sites, probably with
cross-linguistic variation. Furthermore, it should be possi-
ble to independently lexicalize more than one specifier posi-
tion within the (old) IP domain, as has recently been argued
for Icelandic (Jonas & Bobaljik 1993) and Finnish (Holmberg &
Nikanne 1994). An alternative interpretation of the Czech
construction type with expletive and associated thematic
subject might then be that the referential subject occupies
Spec-TP. Further research will have to show whether or not
this is a viable option. It is far from incontravertible
that all the different functional projections of V really
exist. It would be convenient to prove that there is actual-
ly some "use" for the different specifier positions, such as
Spec-TP or Spec-NegP.

5.2.2. Origin and meaning of Czech expletives

In the previous section I argued that expletive pronouns in Czech occupy the structural subject position. We have thus found a counterexample to Hyams's (1986) prediction that languages which use referential pronouns for functional rather than grammatical reasons (i.e. null-subject languages) do not have overt expletive pronouns. Czech is a canonical null-subject language which nevertheless uses lexical expletives.

However, it needs to be mentioned that most of the Czech constructions without expletives are not entirely synonymous with those which do have expletives. In fact, there seem to be three different types of expletive constructions.

The first group includes meteorological verbs and impersonal passives, as illustrated in (13) and (14) above. Here, an overt expletive is invariably associated with an emotive involvement of the speaker, and always implies some sort of emotion such as surprise, joy, disappointment, etc. Above I analyzed *vono* type expletives as structural subjects, despite the fact that some speakers find the co-occurrence of such expletives with material in C awkward and most speakers even reject their co-occurrence with question words in CP. It seems to me that such incompatibilities can be attributed to pragmatic rather than syntactic considerations. A sentence such as *Vono prší* expresses surprise on the part of the speaker about the fact that it is raining - something like the

English 'Wow it's raining!'. It would not make much sense to turn this into a question *'When, wow, is it raining?'.

The same native speakers consulted about the first type of construction also agreed that the expletive in the subject doubling construction in examples (25-32) does not necessarily contribute the same kind of emotive emphasis as the *vono* in the impersonal constructions.[18] Instead, this type of construction seems to be used primarily to lower the style intentionally, depending on the speech situation, although a certain emotional involvement may also pertain. These findings are confirmed by Schalich (1970: 79), who writes about the role of this type of construction in the comedies of the Czech writer Karel Čapek (1890-1938) among others.

> Es [the pronoun] ist hier ein rein aktualisie-rendes Element, das der Äußerung eine persönliche, leicht vertrauliche Note gibt und zugleich die Eindringlichkeit des Gesagten erhöht ... Im authentischen Dialog ist diese Diktion bisweilen durch ein verstärktes Kontaktbedürfnis des Sprechenden motiviert.[19]

---

[18] This seems to be more or less confirmed by other authors. Koenitz (1988) and Hirschová (1984) claim that *vono* type expletives in general have no emotive function. However, their analyses focus primarily on the subject doubling type *Von mu otec všechno dovolí* as in example (28).

[19] 'It [the pronoun] is here purely an actualizing element which lends a personal, slightly intimate touch to the statement and at the same time increases the forcefulness of what has been said ... In authentic dialogue this diction is occasionally motivated by an increased desire for interaction on the part of the speaker.'

Examples (29) and (31) above were generally interpreted as having an additional emotional nuance of warning or threat, as if the overt pronoun *von/vono* expressed a kind of "just you wait" attitude on the part of the speaker. However, the main function of the overt expletive appears to be to mark the discourse as more informal or friendly. It may be exactly the lack of a clear emotive emphasis in most sentences of this type which allows the expletive in this construction type to co-occur with *wh*-words, as illustrated in examples (36) and (37) above.[20]

The third type of construction was shown in the examples under (15). Here, the pronoun *vono* is associated (coindexed) with an extraposed finite clause. It seems to me that in these cases *vono* has actually more of the "pointing" function of a demonstrative and is thus not non-referential as such. In this type of construction *vono* can be replaced with the demonstrative *to* without any change in meaning.

(15') vono/to je možný, že tam nepřijede
      'it is possible that he/she doesn't get there'

The use of pronouns in this position has been reported for other Slavic languages which otherwise do not allow overt expletives. Franks (1990) reports that in Russian the neuter demonstrative *èto* can be used in these contexts. Furthermore,

---

[20]   Most informants also rejected variants of the emotionally colored examples (29) and (31) when material other than the complementizer *protože* preceded the expletive.

a demonstrative pronoun *to* is marginally possible in the same type of construction in Serbo-Croatian and Polish. The fact that these languages use a demonstrative instead of the 3rd person pronoun in this construction indirectly supports my argument about the pointing function of Czech *vono*. The following examples are taken from Franks (1990: 19-21).

(40) (èto) xorošo, čto Ivan pošël domoj   [Ru]
     '(it) is good that I. went home'

(41) ?(to) je dobro, što Jovan ide kući   [SC]
     '(it) is good that J. is going home'

(42) ?(to) jest dobrze, że wszyscy poszli do domu [PO]
     '(it) is good that everybody went home'

An optional use of the pronoun *es* can also be observed in German. As will be discussed in more detail in 5.3.3. German has obligatorily overt quasi-argumental subjects (*weil es/*pro regnet*) and obligatorily null non-argumental subjects (*weil pro/*es hier getanzt wird*). However, in the context of an extraposed finite clause this complementary distribution seems to break down. Both the variants with and without *es* are possible, as shown in the sentences below.

(43) ... weil möglich ist, daß er morgen kommt
     ... weil es möglich ist, daß er morgen kommt
     ...'because (it) is possible that he comes tomorrow'

(44) ... weil schön ist, daß du morgen kommst
     ... weil es schön ist, daß du morgen kommst
     ...'because (it) is nice that you will come tomorrow'

I thus argue that the occurrence of *vono* in the construction type with an extraposed finite clause represents a different phenomenon which is not immediately related to the possibility of overt expletives in a language. However, taking into account the fact that the acceptability of the constructions in Serbo-Croatian and Polish is at best marginal, the occurrence of "real" expletive subjects in Czech may well affect the use of *vono* in these constructions.[21]

The functions of the expletive in the first and second type of construction are related in that they express the involvement/attitude of the speaker. This sheds doubt on Hyams's (1986) position that non-referential elements (i.e.

---

[21]   Franks (1990) treats the pronouns in constructions with extraposed finite clauses as "true" expletives. Recall that he derives the null-subject properties of a language from its nominative case assignment properties (Nominative case must be realized? [OESP]; Nominative case cannot be assigned at LF? [OSP]) Null expletives are generally irrelevant for the LF-visibility of arguments. If a language does not obligatorily assign nominative case, expletives in that language do not get case and thus cannot be lexicalized. However, Franks predicts that a certain type of overt expletive should be possible even in [-OSP,-OESP] languages. Clausal and prepositional complements also need case for their theta-roles to be visible at LF. While infinitivals and PPs have intrinsic case, finite complement clauses must form a chain with the subject position which receives nominative Case. The expletive in the theta-marked chain with a finite clause will thus receive case as well. At this point the expletive pronoun is like any referential pronoun which must have case for LF-visibility but does not need to be lexicalized due to the [-OSP] setting. It may, however, be lexicalized.
Although not explicitly mentioned by Franks (1990), the difference in the acceptability of such elements in Czech on the one hand, and Polish and Serbo-Croatian on the other, might be attributed to the fact that Czech is [-OSP,+OESP] and thus allows optional expletives in general.

expletives) cannot be used for functional/pragmatic purposes. Instead, I claim that overt expletives in Czech should be subsumed under the generally possible occurrence of overt subject pronouns in **marked** discourse. Just as with referential subject pronouns, the use of an overt expletive is evoked by functional considerations. One might interpret lexicalized subject pronouns in Spec-IP (i.e. [Spec,AGRP]) in null-subject languages as an overt realization of the pronominal features of Agr to specially **mark** the discourse.

The occurrence of overt expletives in Czech can thus be shown to be compatible with the avoid pronoun principle. Recall that the use of overt referential pronouns did not contradict the principle, as long as they were used in emphatic contexts. Similarly, by the avoid pronoun principle, a language that licenses null subjects will never use overt expletives *unless* they convey some special pragmatic meaning. Within the scope of this work I will not attempt to analyze the different shades and nuances of this pragmatic meaning. In the following I will use "emphatic expletives" as a general term. Spoken Czech is special in having such expletives. The question as to why expletives in Czech, in contrast to other null-subject languages such as Italian or Spanish, are not really meaningless will be discussed below.

The table below is an "updated" description of possible language types within the licensing/identification framework (cf. also Table 1). It should be emphasized that the following

combinations are **not** parameter settings, but instead descrip-
tions of possible language types derived from the settings for
the actual parameters [*pro* licensed] and [*pro* identified].[22]

Table 3:   Possible Language Types (III)

    i: canonical *pro*-drop

        A [±T,-E]   e.g. Italian
        B [±T,±E]   emphatic expletives, e.g. Czech

    ii: mixed languages

        C [+T,-E]   e.g. Russian, German[23]

        D [+T,±E]   e.g. ??

    iii: canonical non-*pro*-drop

        E [+T,+E]   e.g. English

Types A and B are canonical null-subject languages in the
sense that *pro* is licensed and identified. Both type C and D
belong to the mixed type, i.e. *pro* is licensed but not identi-

---

[22]   I take '+' to mean obligatorily overt, '-'obligato-
rily null and '±' optionally overt. 'T' stands for 'thematic
subjects', 'E' for 'expletive subjects'.

[23]   Although both German and Russian are mixed *pro*-drop
languages, they are still different with respect to their
null-subject properties. In German, quasi-argumental sub-
jects have to be lexicalized, while in standard Russian they
must be *pro*, as illustrated in the contrast between (i) and
(ii).

    (i)   ...weil **es**/*pro* hier kalt ist
           'because it is cold here'

    (ii) ...potomu čto *ono/pro* zdes' xolodno

As mentioned in the introduction, quasi-argumental subjects
share properties of both referential subjects and expletive
subjects, so that such a cross-linguistic "vacillation" with
respect to overtness or covertness should not be too surpris-
ing.

fied. In contrast to type C, type D would be a mixed *pro*-drop language with emphatic expletives. As of yet, this type has not been attested, although it is theoretically possible.[24]

The discussion thus far has shown that the Czech data can actually be accounted for by the licensing/identification framework and that the occurrence of optionally overt expletives in a language which licenses *pro* does not necessarily constitute a violation of Chomsky's avoid pronoun principle.

At this point it seems appropriate to point out some similarities and differences between the predictions made by the (modified) licensing/identification framework summarized in Table 3 and Franks's case-assignment based approach discussed earlier in this chapter (cf. Table 2). Both systems predict the existence of the Czech type, i.e. optionally overt expletives in a canonical null-subject language. According to an analysis à la Franks, spoken Czech has a different parameter setting than a null-subject language which does not allow overt expletives. Within a licensing/identification approach, there is no difference in the parameter settings, and spoken Czech is unusual in that it has emphatic expletives. So far, only very few null-subject languages with overt expletives have been documented. For Franks, the question is why the setting [+OESP,-OSP] seems

---

[24] Woolhiser (1994) mentions the occurrence of expletive pronouns in Ukrainian, Belorussian and Russian dialects. In Chapter 3 I analyzed Russian as a mixed *pro*-drop language. More research is necessary to determine whether or not these dialects may in fact be examples of language type D.

to be so rare.[25] For my analysis, the exceptional status of Czech is justified by the fact that languages normally do not develop emphatic expletives. Franks's system, based on two binary parameters, predicts the existence of only four language types (cf. Table 2). Franks (1990: 14) mentions explicitly that the configuration which is predicted to be **impossible** under his approach is a language with obligatorily overt thematic subjects and optionally overt expletives, which corresponds to type D in Table 3. The existence of mixed *pro*-drop languages with overt expletives (which have not been attested yet) would thus disprove the case-assignment analysis. For Franks, the justification for overt expletives in a language is a setting of [+OESP], and it depends on the setting for the [OSP] parameter whether the expletives are obligatorily ([+OSP]) or optionally ([-OSP]) overt. Optionally overt expletives derive from the setting [-OSP,+OESP]. On the other hand, for a language to have obligatorily overt thematic pronouns, the setting [+OSP] (nominative assignment cannot take place at LF) is assumed, which leads to a contradiction.

Expletives by definition lack any referential properties. How is it then possible that such an element can carry a pragmatically marked reading? The following discussion and exam-

---

[25] Unless, of course, we assume that even in this language type the overtness of expletives is conditioned by functional considerations. It could then be possible that there are many languages with the setting [-OSP,+OESP] which simply could not be differentiated from [-OSP,-OESP] languages due to their lack of emphatic expletives.

ples are based on Trávníček (1962: 16-30), who gives a plausible historical account of the use of expletive pronouns in Czech. He argues that the emotive/emphatic meaning developed out of the *upozorňovací citoslovce* ('attention interjection') *ono/ano*, which corresponds to modern Czech *hle* or *aj* ('look'; 'behold'). In older Biblical texts *ono/ano* was originally used as one of the possible translations of Latin *ecce*. This element had sentential value on its own and was therefore positioned 'outside" the CP of the following clause. The Czech example in (45) illustrates this.

(45) James: 5:7

> **ano**, těžař čaká drahého ővoce (evangeliář Olomoucký,
>   14th century; cited by Trávniček (1962: 20))
>
> **ecce** agricola expectat pretiosum fructum
>
> 'see, the farmer is waiting for the precious fruit'
>
> **aj**, oráč očekává drahého užitku zemského
>   (Kralická Bible, 1613 edition)

Later, this element lost its original function and meaning. The development of its various emotional nuances was accompanied by the loss of its sentential value and incorporation into the sentence structure proper of the following sentence. The original intonational break between the interjection and the next sentence was lost.[26] According to Trávníček,

---

[26]   More evidence for the change in the status of *ano/ono* comes from the placement of clitics. Recall that Czech has second position clitics. Trávníček (1962: 21) writes:

agreeing forms, which in our terms reflect a movement "down the syntactic tree" into subject position, started to develop in the late 15th century. Examples (47) and (48) date back to the first half of the 16th century. There are also examples of non-agreeing *ano/ono* during the transitionary period, as illustrated in example (46) which dates back to the second half of the 14th century.[27]

(46) **ano** však i **ta sama nebesa** vyznávají divy i
svatost jeho
'even the heavens confess his miracles and sainthood'

(47) a **oni**pl **Poláci**pl volili jsú sobě za kníže mladence
jednoho
'and the Poles elected a young man as their ruler'

Když citoslovce ztratilo větnou platnost, vyslovovalo se splývavě s následující větou a cítilo se za její člen, začaly se klást příklonky hned za ně.
'When the interjection lost its sentential value, merged with the following sentence and was felt as its member, the clitics began to attach immediately following it [the original interjection, ML].'

Unfortunately, no examples of clitics and "independent" *ano/ono* are available at this point, so that example (i) is a theoretical reconstruction based on (ii) which is given by Trávníček (*se* is a reflexive clitic).

(i) ano, již **se** setmievá
'behold, it's already getting dark'

(ii) ano **se** již setmievá

[27] These examples are cited from Trávníček. The difference between the *ano* in (46) and (45) is somewhat difficult to prove. Trávníček bases his judgement on the lack of the comma in the former. A comma seems to be consistently present in the older examples where *ano* is still an independent *citoslovečná věta* (interjectional sentence).

(48) oni$_{pl}$ všichni čerti$_{pl}$ vyšli sú z pekla pro tu duši
     'all devils came out of hell for this soul'

In modern Czech agreement is more or less obligatory, unless of course there is no subject with which to agree, which produces the *vono prší/vono se tu tancuje* type.[28] However, some modern dialects may have preserved an intermediate stage of this development with a non-agreeing form of the expletive. Trávníček cites, among others, the Czecho-Moravian dialects as an example, as shown in (49).

(49) vono   často   lidi   viděli neco
     expl.$_{neut.sg}$ often   people$_{pl}$ saw   something
     'people often saw something'

Trávníček (1962: 25) calls the pronoun which evolved out of the interjection *naladové on* ('moody it') and maintains that it expresses the *citové zaujetí* ('emotive bias') of the speaker whenever it is used. While this is true for the *vono prší/vono se pracuje* type, it contradicts my findings somewhat as far as the double subject type is concerned. Recall that

---

[28]   I can only speculate about why this type of construction is not part of the modern *spisovná čeština*. One important source for the codification of the literary language in the early 19th century was the *Kralice Bible* from 1593, which was based on the Moravian dialect. (Kralicy is a town in Moravia). I did not find a single example of subject doubling constructions in the Gospel according to St. Matthew as translated in the *Kralice Bible*, which leads me to conclude that this type of construction in general was not typical for this Bible translation. Recall that the phenomenon of overt expletives was discussed as a feature of the spoken language (*obecná čeština* or *běžně mluvená čeština*) of Bohemia.

most speakers agreed that the expletives in examples (25-32) above do not necessarily contribute an emotive emphasis. However, it is possible that the *náladový odstín* ('emotional nuance') has further changed in the thirty years since Trávníček wrote his book.

As for the use of *vono* in constructions with extraposed finite clauses, I speculate that this function dates back directly to the original common Slavic demonstrative pronoun *onъ, ona, ono*. Common Slavic had no 3rd person pronouns, and it was this demonstrative which developed into 3rd person pronouns in most Slavic languages. Czech is unusual in that it seems to have had an additional source for the 3rd person pronoun, namely the *upozorňovací citoslovce* mentioned above.[29]

5.3. Upper Sorbian

In Chapter 4 I argued that Upper Sorbian does not behave like a canonical null-subject language with respect to the occurrence of referential null pronouns. Sorbian thus produces a serious challenge for a licensing/identification approach to null-subject phenomena. In this section I will investigate whether or not the licensing/identification theory, especially

---

[29]   Trávínček seems to imply that all 3rd person pronouns are derived from the original interjection *ono/ano* (which in turn is most likely derived from the Common Slavic demonstrative pronoun *ono*). He maintains that the double subject constructions occurred at a time before the appearance of normal 3rd person pronouns and basically paved the way for their introduction into the language.

the morphological uniformity hypothesis, makes correct predictions about non-referential subject pronouns in Sorbian.

5.3.1. Syntactic status of overt expletives

The Sorbian verbal paradigm fulfills the morphological uniformity criterion, and thus we would expect *pro* to be licensed. In keeping with Chomsky's avoid pronoun principle and based on the discussion in section 5.2.2. we would then expect that the language should not use overt expletive subjects unless they are able to fulfill some sort of pragmatic function, as we have seen in spoken Czech.

Overt non-referential pronouns are widely used in spoken Sorbian, although they are more or less absent from the literary language. Schuster-Šewc (1974) was, to my knowledge, the first linguist to attempt to give a systematic account of this phenomenon. A discussion of several aspects of his article will be included in this section.

The different forms of the expletive are *wono* and *wone*. These forms are distributed according to dialectal areas. *Wono* is typical for Lower Sorbian, the transitionary dialects (Muskau, Schleife, Spreewitz, Hoyerswerda), and the Northern dialects of Upper Sorbian (Nochten, Reichwalde, Klitten, Kreba), and *wone* for most other Upper Sorbian dialects.[30] In

---

[30]   Schuster-Šewc (1974) points out that the form *wone* is a secondary form derived from *wono*, probably under the influence of the distinction between *to* and *te* for the nominal and neuter attributive use of the demonstrative pronoun.

the South-Eastern tip of Lusatia the form *wón*, which is homonymous with the 3rd person masculine form, is used as a non-referential pronoun instead of *wone*. This phenomenon seems to be strictly limited to a very small area, and speakers from other dialect zones reject it outright. The examples (50-57) below were collected in the "Kerngebiet".

(50) Jana je prajiła, zo chce wona rady kupać hić.
     'J. said, that she would like to go swimmming.'
     Wone pak   je hišće přezyma.
     it   however is still too-cold
     'However it is still too cold'

(51) **wone** je škoda, zo   wječor njepřińdźeš
     it   is pity   that evening not-will-come$_{2sg}$
     'it is too bad that you won't come in the evening'

(52) **wone** njeje rjenje, zo   je přeco hišće tak zyma
     it   not-is nice   that is always still so cold

     'it is not nice that it is still so cold'

(53) poprawom smy   chcyli   dźensa pućować, ale **wone**
     at-first aux$_{1pl}$ wanted today   hike   but it
     so   nažel   dešćuje
     refl unfortunately rains
     'originally we wanted to go for a walk today, but unfortunately it is raining'

(54) **wone** je móžno,   zo   wón hišće přińdźe
     it   is possible that he   still will-come
     'it is possible that he will still come'

     (i)   to ma chwile   'this had time'
     (ii) te blido   'this table'
Wone first appears in Swětlik's 1690 Bible translation.

(55) My hakle jara pozdźe do wjeski přińdźechmy,
we only very late   to village came

a **wone** běše hižo   ćma. **Wone** njeběše nam cyle derje.
and it was already dark it not-was to-us quite good

'We came into the village very late and it was
already dark. It was a little scary.'

(56) dźensa so prózdniny započnu; **wone** je tež wulki čas
today refl vacation start   it   is also high time

'today the break starts; it is also high time'

(57) wšitcy běchu wjeseli; **wone** je so   wjele smjeło,
all   were happy   it   was refl much laughed

rejwało a   spěwało
danced and sung

'everybody was happy; there was a lot of laughing,
dancing, and singing'

Consultants with some degree of formal language training, such
as teachers, ministers, and clerical workers, claimed not to
use this sort of construction, but did not deny familiarity
with the phenomenon. Prof. Helmut Faßke of the sorbisches
Institut analyzed native speaker recordings of spontaneous
speech from a large variety of dialects in preparation for the
syntax volume of the Sorbischer Sprachatlas (Volume XV to
appear) and kindly shared some of his data with me. The fol-
lowing examples are taken from these recordings.[31]

(58) **wono** jo se tak wjele šło, žo šycko spódy wódy stoj
'it rained so much that everything was under water'

---

[31]   Some dialect forms cited do not exist in the stan-
dard language. For further examples of this type see also
Schuster-Šewc (1974).

(59) **wono** było žon jare pozdźe
'it was already very late'

(60) pon bu **wone** wordowało mudrowane a powědane
'then there was a lot of smart talking'

(61) **wone** je dźensa jara mało widźeć
'there is very little to be seen today'

(62) ale **wone** běše nuzne, zo so ta rěka zahaći
'but it was necessary to dam up that river'

(63) **wone** dyrbješe čisće znowa započane wordować
'it had to be started all over again'

(64) ale **wone** tež skoro zas tak je kaž pojeda
'but it is also almost as they say'

(65) **wón** bě to pon tak daloko přišło, zo ....
'it went so far that ...'

Based on the discussion in section 5.2.1., the pronouns in examples (51), (52), (54), and (62) should perhaps not be interpreted as "pure" non-referential pronouns, and instead be seen as being similar to the Czech *vono* co-occurring with extraposed finite clauses. In the following I will concentrate on the other environments in which the expletive-like elements seem to occur.

The use of non-referential pronouns is optional. In clear contrast to Czech, no emotive/emphatic component is associated with the use of these expletives. Thus, if they turned out to be "real" subjects this would be in conflict with the theoretical predictions outlined above. In particular, we would have found a language which lexicalizes non-

emphatic expletive subjects, although *pro* is licensed.

However, in contrast to Czech, there seems to be a syntactic incompatibility between expletives and *wh*-phrases or complementizers. This hypothesis is supported by evidence from various sources. Faßke's corpus of data, for example, provides over 50 examples of overt expletives in many different dialect areas. However, only a single case can be found where *wone* was used after a complementizer. I also asked informants to translate sentences containing overt expletives from German into Sorbian. There was not a single instance in which speakers used an overt expletive after a complementizer or *wh*-word, even if there was an *es* in the German original.[32] The sentences below, produced by an informant who tended to use overt expletives very frequently, illustrate this.

> (66) wone pak   je hišće přezyma
>   it   however is still too-cold
>   'it is still too cold'
>   (es ist aber noch zu kalt)

> (67) mi je žel, zo přeco hišće tak zyma je
>   me is sorry that always still so cold is
>   'I'm sorry that it is still so cold'
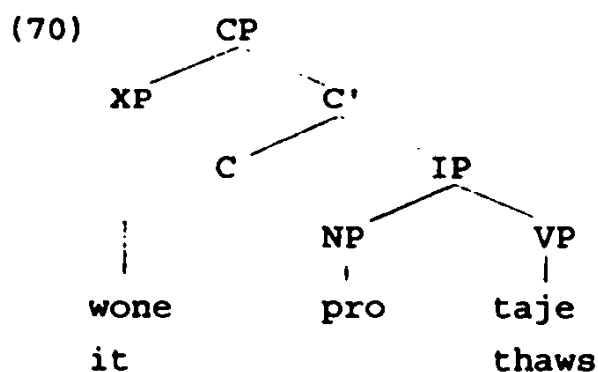>   (mir tut es leid, daß es immer noch so kalt ist)

All Sorbian speakers rejected constructed examples where *wone/wono* was used after a *wh*-word or complementizer. This is shown in examples (68) and (69).

---

[32] For further comparison between Sorbian and German see section 5.3.3.

-183-

(68) hdy (*wone) je zyma
'when is it cold'

(69) dźěći su zrudne, dokelž (*wone) hižo taje
'the kids are sad because it is already thawing'

In contrast to Czech, in Sorbian there is a very clear
syntactic incompatibility between overt expletives and overt
material in the head or specifier of CP. In view of the lack
of any sort of emotive component inherent to these pronouns,
it is not an option to attribute these distributional restric-
tions to pragmatic reasons. This leads me to conclude that
Sorbian expletives are positioned outside [Spec,IP], most
likely in [Spec,CP]. Presumably, *pro* is occupying [Spec,IP]
even if an overt, non-referential *wone/wono* is present, giving
the structure (70) below.

(70)

```
          CP
         /   \
       XP     C'
             /   \
            C     IP
            |    /  \
            |  NP    VP
            |  |     |
          wone pro  taje
          it        thaws
```

To continue this line of reasoning it is useful to look
at the two different types of conjunctions. Subordinating
conjunctions such as *dokelž* ('because') in example (71) reside
in the complementizer position C and take a subordinate clause
(IP) as complement. The expletive *wone*, occupying [Spec,CP]

-184-

lies outside this IP, and a complementizer placed in front of
wone would thus lead to branch crossing and therefore violate
X-bar structure.[33]   Coordinating conjunctions such as *ale*
('but'), on the other hand, conjoin two elements of identical
category, such as two CPs and should therefore not inhibit the
occurrence of the expletive *wone*. This expectation is borne
out by the next pair of sentences, produced by the same infor-
mant (sentence (53) is repeated as (72)).

(71) wón njeje   kopańcu hrać šoł, **dokelž**
      he   neg-aux$_{3sg}$ soccer play went because
      je   so   dešćowało
      aux$_{3sg}$ refl rained
      'he didn't go to play soccer, because it was
      raining'

(72) poprawom smy chcyli dźensa pućować hić, **ale wone**
      so nažel dešćuje
      'originally we wanted to go for a walk, but it is
      unfortunately raining'

A curious type of construction was discovered in a dia-
lect text from the Mješicy region, cited by Faßke and Michałk
(1989: 41). Example (73) is an arbitrary plural construction

---

[33]   The sequence 'wone dokelž', although theoretically
conceivable in terms of the tree structure, is not possible.
Also compare the ungrammaticality of the German sentences
below:
(i)   *weil es getanzt wird
      'because it is being danced'
(ii) *es weil getanzt wird

with a 3rd person singular expletive in initial position.[34]
This indirectly supports my hypothesis that the overt exple-
tive in Upper Sorbian is not in [Spec,IP] ([Spec,AGRP]), since,
otherwise, we would be faced with a very unusual situation of
non-agreement between the singular pronominal subject *wón* and
the plural verb *su pójdali*.

> (73) **wón** su   jow wele pójdali, zo
>   expl.$_{sg}$ aux$_{pl}$ here much said$_{pl}$ that
>   'one used to say here often that'

An analogous example was found among the tapes analyzed by
Faßke, cited here as (74).

> (74) **wón**   to takle činjachu
>   expl.$_{sg}$ that so   did$_{pl}$
>   'it used to be done that way'

5.3.2. Historical perspective

It follows from Seiler's (1830: 113-114) description of the
phenomenon that the situation in the early 19th century seems
to have been more or less the same as in today's spoken lan-
guage.

> Das unbestimmte Pronomen und unbekannte Subjekt *Es*
> wird durch *wono* oder *wone* gegeben, welches dann das
> erste Wort zu sein pflegt: *wono rima*, es donnert;
> *wono s dešćom a se sněhom dźe*, es regnet und schneit
> zusammen; *wone zwoni*, es läutet; *wone so wujasnjuje*,

---

[34]   Mješicy (Meschwitz) is in the Sorbian speaking
territory which uses the form *wón* as an expletive pronoun
(see 5.3.1.).

es hellt sich auf. Da aber schon die dritte Person
singul. neutr. des Zeitworts das *Es* bezeichnet, so
wird in Sätzen, wo Konjunktionen oder andere Rede-
theile voranstehen können, *wono* weggelassen: hdyž
rima, wenn es donnert.[35]

Another indication that these elements were widely used
in the literary language is Smoler's (1841) *Gesprächsbuch.* J.
A. Smoler was one of the leading figures of the *serbske
narodne wozrodźenje* in the 19th century and was particularly
concerned about the creation of a modern, unified Sorbian
literary language. The following are some examples found in
Smoler (1841: 12) (examples are given in modern spelling).

(75) wone so błyska   es blitzt

     wone so počina hić   es fängt an zu regnen

     wone (so) próši   es staubt

     wone (so) kapje   es tröpfelt

     wone so lije   es gießt

Compare this to the *Gesprächsteil* in Kral's (1895) grammar.

(76) chłódno je   es ist kühl

     hižo překapuje   es fängt an zu regnen

     je suche, mokre   es ist trocken, naß

---

[35] 'The indeterminate pronoun and unknown subject *Es*
('it') is given as *wono* or *wone*, which is then the first word
in the sentence: wono rima, es donnert ('it is thundering');
won s deščom a se snĕhom dźe, es regnet und schneit zusammen
('it is raining and snowing at the same time'); wone zwoni,
es läutet ('it is ringing'); wone so wujasnuje, es hellt sich
auf ('it is getting lighter'). However, since the 3rd person
of the verb already shows the *Es* ('it'), *wono* is omitted in
sentences with preceding conjunctions and other parts of
speech: hdyž rima, wenn es donnert ('when it is
thundering').'

je jara ćopło   es ist sehr warm

přeco hiščo mjerznje   es friert immer noch

The use of *wone* thus seems to have disappeared from the literary language in the second half of the 19th century. It is virtually non-existent in today's written language. It seems fair to assume that the puristic movement associated with the *serbske narodne wozrodźenje* artificially caused this sudden disappearance of overt non-referential pronouns from the literary language. Schuster-Šewc comes to a very similar conclusion when he writes (1974: 343):

> Das Meiden des fiktiven Subjekts *wono(e)* in der neueren os. Schriftsprache dürfte zweifellos mit der seit Mitte des vergangenen Jahrhunderts verstärkt auftretenden Tendenz verbunden sein, die Schriftsprache stärker von der Volkssprache abzusetzen und damit auch fremde (d.i. deutsche) sprachliche Einflüsse zurückzudrängen.[36]

According to Schuster-Šewc, overt expletive pronouns appear after 1850 sporadically in those literary works which intentionally reflect the *Volkssprache*. As a typical example he cites the writer A. Černý, whose legends about *Mythiske bytosće łužiskich Serbow* ('Mythical Beings of the Lusatian Sorbs') were published between 1890-95. This example may not be the most appropriate; Černý was Czech by birth and might

---

[36]   'The avoidance of the fictitious subject *wono(e)* in the newer Upper Sorbian written language is undoubtedly connected to the strong tendency, starting in the middle of the last century, to separate the written language further from the vernacular language and thereby to push back alien (i.e. German) linguistic influences.'

have introduced some linguistic "interference" with respect to the use of overt expletive pronouns. However, this should not affect the general validity of Schuster-Šewc's statement.

Schuster-Šewc (1974) points out that in some Old Sorbian manuscripts there are examples of expletives which co-occur with referential subjects in the same sentence. Often they agree in pronominal features with the referential subject, making this phenomenon very reminiscent of the spoken Czech construction discussed above. The most typical text for this is Swětlik's (1690) translation of parts of the New Testament, from which the following examples are taken.

(77) Matthew 8.26

..., ha wona_{fem} so sta jena wulka ćichota_{fem}
and/expl./refl./happened/one/great/calmness

(78) Matthew 8.32 (cited from Michałk (1972: 93)

ha hlejćo, wona_{fem} z hujom woteńdźe ta cyła
črjóda_{fem} přez nahły spad do toho morja nutř
and/behold/expl./with/quickness/went-away/
that/whole/herd/over/steep/fall/to/the/sea/inside

It is interesting to compare Swětlik's 1690 translation of various Biblical passages with the 1750 version of the same passages. The latter, according to Šołta (1921: 33), should be considered as an edited Swětlik translation rather than a completely new one. Surprisingly, the constructions with agreeing expletive-like elements seem to have virtually

disappeared in the 1750 pericopes. The examples below provide
some direct comparisons.

(79) a. Matthew 8:13 (Swětlik 1690)

ha wón$_{masc}$ rjekny Jězus$_{masc}$ k tomu hawtmanej:
... ha wón$_{masc}$ bě tón hólc$_{masc}$ wustrowjeny wo
tejsamej štundźe.
and/expl./said/Jesus/to/the/centurion/
and/expl./was/that/boy/healed/in/same/hour

b. Matthew 8:13 (1750 pericopes)

ha Jězus rjekny tomu hauptmanej: ... ha tón
hólc bó wustrowjeny wo tejsamoj štundźe.

(80) a. Matthew 8:8 (1690)

ha wón$_{masc}$ budźo bóć mój wotroček$_{masc}$ wustrojeny.
and/expl./will/be/my/servant/healed

b. Matthew 8:8 (1750)

ha budźe mój služownik wustrowjeny.

(81) a. Matthew 8:25 (1690)

Ha woni$_{pl}$ přistupichu joho hučowniki$_{pl}$ k njomu.
and/they/stepped-to/his/disciples/to/him

b. Matthew 8:25 (1750)

Ha ći joho wučownicy stupichu k njomu.

(82) a. Matthew 8:3 (1690)

Ha wón$_{masc}$ bě hnydy tón joho wusad$_{masc}$
wušwarnjeny.
and/expl./was/immediately/this/his/leprosy/healed

b. Matthew 8:3 (1750)

Ha hnydy bó ton joho wusad wušwarnjeny.

Since this difference between the two editions could not yet be due to puristic endeavors, it confirms my hypothesis that the construction with agreeing expletives was never really typical for Old Sorbian.[37] Instead, it was more or less part of Swětlik's personal idiolect. Perhaps it was felt to be strange and thus removed by the editors of the 1750 version. One might speculate that Swětlik was particularly influenced by Czech constructions with agreeing expletives since he spent about ten years in Bohemia. However, a comparison with the analogous passages from the *Wenceslaus Bible* (1677) shows that this special feature of Swětlik's language could not be due to direct influence from the Czech Bible. As already mentioned above, there are also no instances of such expletives in the older *Kralická Bible* (1613). It is thus more likely that Swětlik was influenced by spoken Czech.

5.3.3. Upper Sorbian and German

When considering the possible origin of non-referential pronouns in Upper Sorbian, one may consider German influence as a possible source.

(83) **es** regnet; **es** ist kalt; **es** ist dunkel

However, as first illustrated by Schuster-Šewc (1974), the

---

[37] Also recall that there was only a very insignificant decline in the number of overt referential pronouns in the 1750 version compared to Swětlik (1690). For details see 4.4.

syntactic distribution of German es and Sorbian *wone/wono* is notably different. First of all, the occurrence of overt non-referential pronouns in Upper Sorbian is always optional, even in those dialects which strongly prefer the version with the expletive. In German, certain non-referential pronouns, namely those which are subjects of the so-called "weather-verbs", are obligatorily overt, independent of their position in the sentence.

(84) a) **es** regnet morgen   'it is raining'
     b) wann regnet **es**   'when is it raining'
     c) weil **es** regnet   'because it is raining'

As explained above, in Sorbian, the expletive is barred from occurring in the syntactic positions in (84b) and (84c) because of the structure posited in (70). In German, the quasi-argumental *es* which is associated with meteorological verbs can (and in fact must) co-occur with a *wh*-word in [Spec,CP] or a complementizer in position C. These distributional properties indicate that this type of German *es* is indeed the structural subject of the sentence.[38]

_____

[38]   *Es* in these clauses is base-generated in [NP,IP], but may move to a different position at surface structure. The following structures assume that subordinate clauses represent the underlying word order and that root clauses are derived by obligatory movement of the finite verb into C.
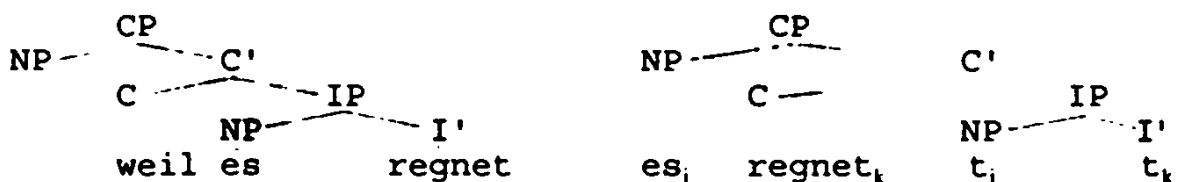
Michalk (1965a, 1977) cites some interesting instances of "Sorbisms" in German. Below are examples where Sorbian native speakers omit the (obligatory) German *es*, cited from Michalk (1965: 713) (# indicates the place where the es is missing).

> (85) und wenn [#] dann soweit ist, dann
>   'and if [it] then is time, then'
>
> (86) sehr kalt ist [#] bei uns nicht mehr.
>   '[it] is not very cold anymore here'
>
> (87) tagsüber taut [#] noch
>   'during the day [it] is still thawing'

There is another type of German *es* which seems not to be a structural subject but resides rather in the same structural position as the Upper Sorbian non-referential *wone*, i.e. in [Spec,CP]. This is illustrated in the impersonal sentences below.[39]

> (88) a. **es** wurde viel getanzt und gesungen
>   'there was a lot of dancing and singing'
>
>   b. ..., weil (***es**) viel getanzt und gesungen wurde
>
>   c. wann wurde (***es**) viel getanzt und gesungen

Examples (88b) and (88c) are ungrammatical for exactly the

```
            CP
XP                C'
       C.                IP
              NP  ..           I'
wann  regnet_i      es              t_i
```

[39]   Cf. also fn. 10.

same reasons that Upper Sorbian *wone* and co-occurring subordi-
nating conjunctions or *wh*-words are ungrammatical. This type
of *es* is not a structural subject and therefore is not gener-
ated in [Spec,IP]. Instead, it seems to be inserted into
[Spec,CP] due to language specific requirements that this
position has to be lexically filled in German declarative root
clauses. As illustrated earlier, sentence (88a) actually has
an empty NP (*pro*) in subject position [NP,IP], which identi-
fies German as a mixed *pro*-drop language. Referential and
quasi-argumental subjects have to be overt; non-argumental
subjects are obligatorily *pro*.

The German "non-subject" *es* and Upper Sorbian non-refer-
ential *wone* seem to serve the same purpose. In terms of func-
tional sentence perspective they occupy the theme position,
presumably to allow the core of the message to be placed to-
ward the end of the sentence. Some Upper Sorbian speakers who
reject the use of overt expletives consistently placed *wonka*
('outside') in sentence initial position, apparently to
achieve the same results. The following examples were elicited
in German.

(89) wonka so dešćuje   - es regnet
      'outside it rains'

(90) wonka je zyma   - es ist kalt
      'outside it's cold'

5.4. Summary

In this chapter I analyzed the occurrence of expletive-like overt pronouns in Czech and Sorbian, which at first glance appear to pose problems for a licensing/identification system for null subjects. According to standard views of null-subject properties, the optional occurrence of overt non-referential pronouns is unexpected in any language where *pro* is licensed. I showed, first, that spoken Czech, despite its status as a canonical null-subject language, allows overt expletive subjects. However, because the use of Czech expletives is based on functional considerations, I argued that this phenomenon does not really contradict theoretical predictions made by the licensing/identification framework in conjunction with the avoid pronoun principle. Sentences with and without overt expletives are usually not totally synonymous, and it thus seems that Czech expletives can convey some special pragmatic meaning. Therefore they should be subsumed under the generally possible occurrence of overt subject pronouns in marked context. I then followed Trávníček's argument about the possible derivation of such "functionally charged" expletives from a sentential interjection with the function of "drawing somebody's attention to something".

In contrast to Czech, no emotive component is associated with the use of expletives in Sorbian. However, the syntactic incompatibility of expletives with *wh*-phrases and complementizers indicates that expletives are not structural subjects.

Instead I argued that these elements occupy [Spec,CP], presumably with *pro* in the subject position, which produces the structure [_{CP} wone [_{IP}*pro* taje]] ('it's thawing'). The positioning of Sorbian *wone* is analogous to and perhaps under the influence of German non-argumental *es* (which crucially does **not** include the subject of weather-verbs) which is inserted in Verb-second contexts and **only** legitimate in that context. In contrast to Sorbian, German has another type of non-referential *es*, the so-called quasi-argumental *es*, which is a real subject and must always be overt. This type of *es* is the subject of meteorological verbs in German. It follows then that in Sorbian, there is no difference in the position of *wone* in the following examples, while in German there is a different position of *es*.

> (91) a. wone taje
>
> 'it is thawing'
>
> [_{CP} wone [_{IP}*pro* taje]]
>
> wone je so wjele smjało
>
> 'there was a lot of laughing going on'
>
> [_{CP} wone je so [_{IP}*pro* wjele smjało]]
>
> (92) a. es taut
>
> [_{IP} es taut]]
>
> b. es wurde viel gelacht
>
> [_{CP} es wurde [_{IP}*pro* viel gelacht]]

6. CONCLUSION

In this work I discussed data from Russian, Sorbian, and Czech which appear to challenge licensing/identification systems such as the one suggested by Jaeggli/Safir (1989). The application of a set of three null-subject diagnostics to Slavic languages in Chapter 3 suggested that most West and South Slavic languages are canonical *pro*-drop languages, while Russian, taken to be representative of East Slavic, is not. After suggesting an intuitive explanation for this split, I set out to formalize the difference between these two groups in terms of the inconsistent realization of the functional projection AGR_{pers} in Russian.

In Chapter 4 I showed that Sorbian does not possess all the properties of a canonical null-subject language, despite claims to the contrary by language purists. To explain and illustrate this puristic error I cited numerous examples from the written language (drawn both from early and contemporary texts) and the spoken languages. This was followed by a critical discussion of three possible answers to the question of why Sorbian is not a canonical null-subject language. Within a licensing/identification system it would be necessary to answer the question of why the identification of thematic null subjects in Sorbian is blocked despite the obviously 'rich' verbal morphology. Two (more or less successful) models were then formulated. One was derived from the possibility that the number of syncretisms allowable for successful identifica-

tion had been exceeded, and the other was based on word order, specifically the verb-final preference found in Sorbian, which is peculiar in Slavic. However, both of these models presented potential difficulties in the handling of certain data, or produced counter-intuitive predictions. In the final analysis it was determined that the licensing/identification system in its restricted form cannot account for the Sorbian data. As an alternative, I proposed a model in which Sorbian is analyzed as a truly optional *pro*-drop language in which *pro* is licensed and identified, and where the presence of atonic personal pronouns causes the null and overt variants to be in competition.

In chapter 5 I discussed the occurrence of overt expletives in Czech and Sorbian. This analysis led to the conclusion that these elements, though different in their syntactic status, are actually not a challenge for the predictions of possible language types made by the kind of identification and licensing system suggested by Jaeggli/Safir (1989) in conjunction with the avoid pronoun principle.

BIBLIOGRAPHY

Avanesov R.I. and V.V. Ivanov. (1982) *Istoričeskaja grammatika russkogo jazyka*. Moskva: Nauka.

' Babby, L. (1989) "Subjectlessness, external subcategorization, and the Projection Principle". *Zbornik za filologiju i lingvistika* 32: 7-39.

Bauer, J. (1968) "Strukturní typy slovanské věty a jejich vývoj". *Československé přednášky pro VI. mezinárodní sjezd slavistů*. Praha: Akademia.

Benedicto, E. (1994) "Agr, θ-features and V-movement: Identifying *pro*". *University of Massachusetts occasional papers in linguistics* 17: 1-18.

Bennis, H. and L. Haegemann. (1984) "On the status of agreement and relative clauses in West-Flemish". De Geest and Putseys eds. *Sentential complementation*. Dordrecht: Foris, 33-53.

Berneker, E. (1900) *Die Wortfolge in den slavischen Sprachen*. Berlin: Behr's Verlag.

Bloomfield, L. (1935/1958) *Language*. New York:Holt and Co.

Borer, H. (1989) "Anaphoric AGR". Jaeggli O. and K.J. Safir eds. *The null subject parameter*. Dordrecht: Kluwer, 69-109.

' Bouchard, D. (1989) *On the content of empty categories*. Dordrecht: Foris.

Cardinaletti, A. (1990) "*Es pro* and sentential arguments in German". *Linguistische Berichte* 126: 135-164.

Carlton, T. (1990) *Introduction to the phonological history of the Slavic languages*. Columbus,OH: Slavica.

Chomsky, N. (1981) *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1982) *Some concepts and consequences of the theory of government and binding*. Cambridge,MA: MIT Press.

Chomsky, N. (1986) *Barriers*. Cambridge,MA: MIT Press.

Černyx, P.J. (1954) *Istoričeskaja grammatika russkogo jazyka*. Moskva: Nauka.

De Bray, R. (1963) *Guide to the Slavonic languages*. London: Dent & Sons.

Elle, L. (1991) "Die Sorben in der Statistik". *Die Sorben in Deutschland*. Bautzen: Maćica Serbska, 21-25.

Elle, L. (1992) "Zur Sprachsituation der Lausitzer Sorben". *Slavia Occidentalis* 48/49: 67-83.

Emonds, J. (1976) *A transformational approach to syntax*. New York: Academic Press.

Faska, H. (1964) "Je personalny pronomen při formje finitneho słowjesa njetrěbny?" *Serbska šula* 64: 572-576.

Faßke, H. (1979) "Semantika a distribucija personalnych pronomenow w hornjoserbskej spisownej rěči". *Lětopis instituta za serbski ludospyt* 26, 1-23.

Faßke, H. (1981) *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie*. (in conjunction with F. Michałk) Bautzen: VEB Domowina-Verlag.

Faßke, H. and F. Michałk (1989) "Dialektny tekst z Mješic". *Lětopis instituta za serbski ludospyt* 36: 38-52.

Feldstein, R. (1987) "*Czyta* vs. *czytaj* and the determination of Polish conjugational desinences" *International journal of Slavic linguistics and poetics* XXXV-XXXVI: 65-76.

Franks, S. (1990) "On the status of null expletives". *Lingua* 81: 1-24.

Franks, S. (1994) "Null subject characteristics of Slavic languages". Paper presented at the annual meeting of the American Association of Teachers of Slavic and Eastern European Languages.

Franks, S. (1995) "Null subject phenomena". *Parameters of Slavic morphosyntax*. Cambridge: Oxford University Press.

Franks, S. and G. Greenberg. (1993) "The functional structure of Slavic clauses". *Formal approaches to Slavic linguistics: The Ann Arbor meeting*. Ann Arbor: Slavica 77-108.

Gribble, C. (1973) *Medieval Slavic texts*. Cambridge,MA: Slavica.

Haegeman, L. (1994) *Introduction to government & binding theory*. Cambridge,MA: Basil Blackwell.

Haider, H. (1991) "PRO-PLEME". *Strukturen und Merkmale syntaktischer Kategorien*. Tübingen: Gunter Narr. (*Studien zur deutschen Grammatik, 39.*)

Hammer, L.B. (1985) *Prague colloquial Czech*. Ph.D. dissertation, Indiana University.

Hirschová, M. (1984) "K jednomu typu výpovědí běžně mluvené češtiny". *Slovo i Slovesnost* 45: 273-275.

Holmberg, A, Nikanne U. (1994) "Expletives and subject positions in Finnish". *Proceedings of NELS* 24: 73-187.

Huang, C.-T. J. (1984) "On the distribution and reference of empty pronouns". *Linguistic inquiry* 15: 531-574.

Hyams, N. (1986) *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

Innis, J. et al. (1985) *Intermediate and advanced Russian reader*. Ann Arbor: Ardis.

Ivanov, V.V. (1980) *Istoričeskaja grammatika russkogo jazyka*. Moskva: Prosveščenie.

Jaeggli, O. (1982) *Topics in Romance syntax*. Dordrecht: Foris.

Jaeggli, O. (1986) "Arbitrary plural pronouns". *Natural language and linguistic theory* 4: 43-76.1

Jaeggli, O. and N. Hyams. (1987) "Morphological uniformity and the setting of the null subject parameter". *Proceeding of NELS* 18: 238-253.

Jaeggli, O. and K. Safir. (1989) *Null subject parameter*. Dordrecht: Kluwer.

Jagić, V. (1900) "Beiträge zur slavischen Syntax". *Denkschriften der Kaiserlichen Akademie der Wissenschaften*. XLVI: 1-88. Wien: Carl Gerold's Sohn.

Jakobson, R. (1939) "Signe Zéro". *Bally, C. ed. Mélanges de Linguistique*. Geneve: Georg et Cie.

Jakobson, R. (1948) "Russian conjugation" *Selected writings* II: 119-129. Mouton: The Hague.

Janaš, P. (1984) *Niedersorbische Grammatik*. Bautzen: VEB Domowina-Verlag.

Jenč, R. (1959) "Městno finitnych formow pomocneho słowjesa a participa wuznamoweho słowjesa w hornjoserbskej sadźe". *Lětopis instituta za serbski ludospyt* 6: 3-47.

Jenč, R. (1964) "K prašenju normy w hornjoserbskej spisownej rěči". *Serbska šula* 17, 1964.4: 247-254.

Jenč, R. (1965) "K prašenju normy w hornjoserbskej spisownej rěči". *Serbska šula* 17, 1965.5: 309-317.

Jentsch, H., Michalk S. 1971. *Sorbische Dialekttexte.* IX Säuritz, Ostro, Jauer, Kreis Kamenz. Bautzen: VEB Domowina-Verlag.

Jonas, D. and J.D. Bobaljik (1993) "Specs for subjects: The role of TP in Icelandic". *MIT working papers in linguistics* 18: 59-98.

Jordan, F.P. (1841) *Grammatik der sorbisch-wendischen Sprache.* Prag: Friedrich Ehrlich.

King, T. (1993) *Configuring topic and focus in Russian.* Ph.D. dissertation, Stanford University.

Koch, J. (1991) *Ha lećala je módra wróna.* Bautzen: VEB Domowina-Verlag.

Koenitz, B. (1988) "Zur Charakteristik der Satzformen mit "expletivem" Personalpronomen im Tschechischen und Deutschen". *Linguistische Arbeitsberichte* 66: 8-14, Karl-Marx-Universität Leipzig.

Koktová, E. (1992) "Anaphoric expressions across clause boundary in pro-drop languages". *Prague bulletin of mathematical linguistics* 58: 39-63.

Kosta, P. (1990) "Zur formalen Lizensierung und inhaltlichen Determination leerer Subjekte im Russischen". *Slawistische Beiträge* 260: 117-165. München: Verlag Otto Sagner.

Kral, G. (1895) *Grammatik der Wendischen Sprache in der Oberlausitz.* Bautzen: Druck und Verlag von Schmaler.

Kroch, A. (1989) "Function and grammar in the history of English: Periphrastic do". *Fasold, R. and D. Schiffrin (eds). Language change and variation.* Amsterdam: Benjamins.

Kuen, H. (1957) "Die Gewohnheit der mehrfachen Bezeichnung des Subjekts in der Romania und die Gründe ihres Aufkommens". *Syntactica und Stilistica.* (Festschrift für Gamillscheg). Tübingen: Max Niemeyer.

Lamprecht, A. et al. (1986) *Historická mluvnice češtiny.* Praha: Státní pedagogické nakladatelství.

Leinonen, M. (1983) "Generic zero subjects in Finnish and Russian". *Scando-Slavica* 29: 143-161.

Liebsch, G. (1884) *Syntax der wendischen Sprache in der Oberlausitz.* 1973 reprint. Bautzen: VEB Domowina-Verlag.

Lindseth, M. (1994) "Expletive and thematic subjects in West Slavic". *Studies in linguistic sciences* 24.2: 333-346.

Lindseth, M. and S. Franks (1996). "Licensing and identification of null subjects in Slavic". In *Formal approaches to Slavic linguistics: The College Park meeting,* 199-225.

Lindseth, M., Ludwig, J. (1992) "The overt, non-referential pronoun in Czech, Lusatian and German". Paper presented at the annual meeting of the American Association of Teachers of Slavic and Eastern European Languages.

Lorenz, K. (ed.) (1981) *Serbska čitanka.* Leipzig: Reclam.

Lötzsch, R. (1968) "Einige Auswirkungen des Purismus auf die grammatische Normierung slawischer Schriftsprachen". *Sorabistiske přinoški k VI. mjezynarodnemu kongresej slawistow w Praze 1968. Spisy instituta za serbski ludospyt* 26. Bautzen: VEB Domowina-Verlag.

Meier, G.F. (1961) *Das Zero-Problem in der Linguistik.* Berlin: Akademie-Verlag.

Michalk, F. (1956-57) "Słowosłěd w Serbščinje". *Lětopis instituta za serbski ludospyt* 17: 1-29.

Michalk, F. (1962a) "Wliw němčiny na położenje finitneho słowesa w serbskej sadźe" *Otázky slovanské syntaxe.* Praha: Statní Pedagogické Nakldatelství.

Michalk, F. (1962b) "Der Einfluß des Deutschen auf die Stellung des Verbum finitum im sorbischen Satz". *Zeitschrift für Slavistik* 62, 232-261.

Michalk, F. (1965a) "Wliw serbščiny na syntaksu němčiny we Łužicy". *Serbska šula* 18: 711-718.

Michalk, F. (1965b) "Enkliza w serbščinje". *Serbska šula* 18: 124-128.

Michalk, F. (1970) "K prašenjam słowosłěda w serbskich dialektach". *Lětopis instituta za serbski ludospyt* 17: 1-29.

Michalk, F. (1972) "Latinizmy a germanizmy w rěči J. H. Swětlika". *Lětopis instituta za serbski ludospyt* 19: 52-106.

Michalk, S. (1977) "Typische Erscheinungen des sorbisch-deutschen Sprachkontakts". *Methodische Beiträge zum Sprachunterricht* 1977: 75-105.

Michalk, S. and H. Protze. (1974) *Studien zur sprachlichen Interferenz. Deutsch-sorbische Dialekttexte aus Radibor, Kreis Bautzen*. Bautzen: VEB Domowina-Verlag.

Miklosich, F. (1883) *Subjectlose Sätze*. Wien: Wilhelm Braumüller.

Montalbelli, M. (1986) "How Pro is it?". Jaeggli, O. and Silva-Corvalan, C. eds. *Studies in Romance linguistics*. Dordrecht: Foris.

Müller, G. (1989) "Eine Geschichte ohne Subjekt. Zur Entwicklung der *pro*-Theorie". *Linguistische Berichte* 119: 3-52.

Nakhimovsky A.D. and R.L. Leed (1987) *Advanced Russian*. Columbus,OH: Slavica Publishers.

Nowak, H. (1988) *Powědamy dolnoserbski. Gutes Niedersorbisch*. Bautzen: VEB Domowina-Verlag.

Orešnik, Janez. (1986) "The obligatorium of unemphatic pronoun subjects in the Germanic languages". *Filologija* 14: 261-269.

Oschlies, W. (1991) *Die Sorben - Slawisches Volk im Osten Deutschlands*. Bonn: Friedrich-Ebert-Stiftung.

Perlmutter, D.M. (1971) *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart and Winston.

Pětrancec, M. (1978) "Bjezsubjektowe werbalne sady w rušćinje a jich přeložk do serbšćiny". Diplomarbeit, Leipzig.

Pollock, J.Y. (1989) "Verb movement, universal grammar, and the structure of IP". *Linguistic inquiry* 20: 365-424.

Prince, E. F. (1993) "On the discourse functions of syntactic form in Yiddish: Expletive ES and subject-postposing".

Goldberg D. ed. *The field of Yiddish*. Evanston: Northwestern University Press.

Rizzi, L. (1982) *Issues in Italian syntax*. Dordrecht: Foris. (Chapter IV, "Negation, *wh*-movement and the null subject parameter")

Rizzi, L. (1986) "Null objects in Italian and the theory of pro". *Linguistic inquiry* 17: 501-557.

Roberge, Y. (1990) *The syntactic recoverability of null arguments*. Kingston: McGill-Queen's University Press.

Roberts, I. (1993) *Verbs and diachronic syntax*. Dordrecht: Kluwer.

Rohrbacher, B. (1995) "Explaining the syntactic consequences of "rich" agreement morphology: On the licensing of V-to-AgrS raising and pro". *Proceedings of WCCFL* 13: 350-364.

Safir, K. (1985) *Syntactic chains*. Cambridge University Press.

Schalich, G. (1970) *Expressive sprachliche Mittel im modernen tschechischen Lustspiel*. München: Fink Verlag.

Schneider, F. (1958) *Grammatik der wendischen Sprache katholischen Dialect's*. Bautzen: Hiecke.

Schuster-Šewc, H. (1967) *Sorbische Sprachdenkmäler: 16.-18. Jahrhundert*. Bautzen: VEB Domowina-Verlag.

Schuster-Šewc, H. (1974) "Sätze mit fiktiven Subjekt vom Typ os. 'Wono so dešćuje / ns. to se pada' und ihre Stellung in der slawischen Syntax". *Zeitschrift für Slawistik* 19: 340-352.

Schuster-Šewc, H. (1976) *Gramatika hornjoserbskeje rěče. Syntaksa*. Bautzen: VEB Domowina-Verlag.

Seiler, A. (1830) *Kurzgefaßte Grammatik der sorbischen Sprache*. (Fotomechanischer Neudruck Bautzen: VEB Domowina-Verlag, 1973).

Short, D. (1994) "Czech". *Comrie B. & G. Corbett eds. The Slavonic languages*. London: Routledge, 455-532.

Siebenschein, H. (1988) *Deutsch-Tschechisches/Tschechisch Deutsches Wörterbuch*. Praha: Státní Pedagogicke Nakladatelství.

Smoler, J.A. (1841) *Maly Serb. Wendisch-deutsche Gespräche*. Bautzen: Wellersche Buchhandlung.

Speas, M. (1994) "Null arguments in a theory of economy of projection". *University of Massachusetts occasional papers in linguistics* 17: 179-208.

Sperling, A. (1990) "Arbitrary pronominals in Slavic". Unpublished manuscript, Indiana University.

Spieß, G. (1995) "Die sprachliche Situation bei den Sorben in der Niederlausitz". Lětopis instituta za serbski ludospyt 42: 59-62, Sonderheft.

Sportiche, D. (1988) "A Theory of floating quantifiers and its corollaries for constituent structure". *Linguistic inquiry* 19: 425-449.

Sprouse, R. and B. Vance (1995) "An Explanation for the decline of null pronouns in certain Germanic and Romance languages". To appear in *Language creation and change: Creolization, diachrony, and development*. MIT Press.

Šolta, B. (1921) "Prěni hornjoserbski přeložk biblije". *Časopis Maćicy Serbskeje*, 3-34.

Šolta, J. et al. (1984) *Nowy biografiski słownik*. Bautzen: VEB Domowina-Verlag.

Stone, G. (1994) "Sorbian". Comrie B. & G. Corbett eds. The Slavonic Languages. London: Routledge, 593-685.

Taraldsen, K.T. (1978) "On the nominative island condition, vacuous application and the *that*-trace filter". Mimeographed ms., MIT, Cambridge; distributed by the Indiana University Linguistics Club, Bloomington, Indiana, 1980.

Tomaselli, A. (1982) "Das unpersönliche "es" - Eine Analyse im Rahmen der generativen Grammatik". *Linguistische Berichte* 102: 171-190.

Townsend, C. (1990) *A description of spoken Prague Czech*. Columbus, OH: Slavica.

Trávníček, F. (1951) *Mluvnice spisovné češtiny* II. Praha: Slovanské Nakladatelství.

Trávníček, F. (1962) *Historická mluvnice česká* III. Praha: Státní Pedagogické Nakladatelství.

Tschernych, P.J. (1957) *Historische Grammatik der russischen Sprache*. Halle: Niemeyer.

Waurick, I. (1968) "Der Inhaltssatz im Niedersorbischen". *Zur grammatischen und lexikalischen Struktur der slawischen Gegenwartssprachen*. Halle: Max Niemeyer, 113-130.

Williams, E. (1987) "Implicit arguments, the binding theory, and control". *Natural language and linguistic theory* 5: 151-180.

Woolhiser, C. (1994) "Expletive pronouns in Slavic: Subject or topic?" Indiana University manuscript.

Zubatý, J. (1917) "Die 'man'-Sätze". *Zeitschrift für vergleichende Sprachforschung* XL: 478-520.

Bibles and Pericopes:

<u>Upper Sorbian:</u> (in chronological order)

S.Matthaeus und S.Marcus, translated from German by M. Frencel, 1670.

Swjate Sćenja, Lekcijony ha Epistle na te Njedźele, ha swjate Dny toho cyłoho Lěta, translated from Latin by J.H. Swětlik, 1690 (pericopes).

Biblija, to je cyłe Swjate Pismo Stareho a Noweho Zakona, translated from German by several Lutheran ministers, 1728.

Swjate sćenja, lekciony a epistle ... 1750 (pericopes)

Biblia, to je cyłe Swjate Pismo Stareho a Noweho Zakona, translated from German by Kühn 1823.

Swjate sćenja, lekciony a epistle ... 1848 (pericopes).

Zowy Zakoń našeho knjeza Jězusa Chrystusa, translated by Jakub Buk, 1862.

Nowy zakón, do hornjosebšćiny po rjedźe Vulgaty přełožištaj, translated form Latin by J. Lusčanski a M. Hórnik, Budyšin 1896.

Perikopy. Epistle, lekcije a sćenja ... 1909 (pericopes).

Sćenje po swjatym Mateju, originally translated by M. Frencel and newly translated by the Bible commission of Lutheran Sorbian ministers, 1960.

Swjate Pismo Noweho Zakonja, 1966.

<u>Lower Sorbian:</u>

Das niedersorbische Testament des Miklawuš Jakubica, 1548. (publ. by H. Schuster-Šewc, Akademie-Verlag: Berlin, 1967.)

Nowy Testament, translated by G. Fabricius 1709.

<u>Czech</u>:

Biblí Svatá aneb všecka svatá písma Starého i Nového Zákona. Podle posledního vydání Kralického z roku 1613. ("Kralická Bible")

Biblí Česká 1677 ("Wenceslaus Bible").

Biblí Česká čili Písmo Svaté Starého i Nového Zákona, 1889.

<u>Latin</u>:

Biblia Sacra, Vulgatae editionis, 1914 ("Latin Vulgate").

<u>German</u>:

Die Bibel oder die ganze Heilige Schrift, translated by M. Luther, 1522.