The History of Null Subjects in North Slavonic

A Corpus-Based Diachronic Investigation

Habilitation Thesis

University of Regensburg

Dr. Roland Meyer

October 2011

Contents

Preface

The present thesis arose as part of a three-years grant from the German Research Founda-
tion (DFG) entitled "Corpus linguistics and diachronic syntax: The grammaticalisation of
peripheral subjects in Slavonic languages" (project leader: Prof. Dr. Björn Hansen, apl. Prof.
Dr. Ernst Hansack), at the University of Regensburg. Without this grant, and without the
readiness of my supervisor and the University of Regensburg to allow me a long leave from
teaching in order to work on my qualification, this thesis could not have been written. I
wish to cordially thank the DFG, my University, and my supervisor for their support. Björn
Hansen was always open-minded and set a lot of confidence in my plans, for which I am espe-
cially grateful. Ernst Hansack was supportive and a valuable source of knowledge on older
stages of the Slavonic languages. I would also like to thank my colleagues on the project,
Christine Grillborzer, Arek Danszczyk, Mirjam Zumstein, Iryna Parkhomenko and Uliana
Yazhinova, for their enthusiasm, their fine work and their flexibility when necessary.

Furthermore, I wish to thank all researchers who genereously granted me access to their
text collections – the Institute of the Czech National Corpus, headed by Prof. F. Čermák and
Dr. V. Schmiedtová, and especially the head of its diachronic section, Prof. K. Kučera, the
team of the *Vokabulář webový* at the AV ČR, mainly Dr. B. Lehečka, Dr. A. Černá, and PD Dr.
K. Oliva; the Institute of Polish language at the Academy of Sciences in Kraków, specifically
prof. Rafał Górski; Dr. Thomas Menzel and Prof. Gerd Hentschel, who allowed us to include
the Göttingen Corpus into PolDi; Prof. V. Baranov for a stimulating exchange of ideas on Old
Russian corpora; Prof. Dag T. T. Haug and Dr. Hanne Eckhoff, who offered me the opportu-
nity to use the PROIEL parallel corpus of Greek and Old Church Slavonic – and, last but not
least, Amir Zeldes and the ANNIS-2 team at the HU Berlin and the University of Potsdam,
for their cooperation. Stefanie Dipper, Ruprecht von Waldenfels and Adam Przepiórkowski
accompanied and commented on parts of the computational work which went into this the-
sis; Uwe Junghanns, Denisa Lenertová, Frank Seifart, Sabine Stoll and Luka Szucsich were
valuable partners in linguistic discussions at various points.

Last but not least, I wish to thank my family, above all Denisa, for their patience and their
practical help when it was needed.

Abbreviations

ACC	accusative		MP	modal particle
ACT	active		N	neutrum
AOR	aorist		NEG	negation
arb	arbitrary		NOM	nominative
AUX	auxiliary		NP	noun phrase
BCS	Bosnian/Croatian/Serbian		NSL	Null Subject Language
CL	clitic		NTF	-no/-to form
CMP	comparative		OCS	Old Church Slavonic
COND	conditional		OES	Old East Slavonic
CP	complementizer phrase		P	person
CS	Conceptual Structure		PAS	passive
CVB	converb		PL	plural
Cz	Czech		Pol	Polish
DAT	dative		POSS	possessive
DP	determiner phrase		PP	prepositional phrase
DU	dual		PRP	prepositive
EPP	Extended Projection Principle		PRS	present
F	feminine		PT	past
Fin	Finnish		PTCL	particle
FUT	future		QU	question marker
GEN	genitive		QVE	quantifying variability effect
Ger	German		REFL	reflexive
IMP	imperative		Ru	Russian
IND	indicative		SBJ	subjunctive
INE	inessive case		SF	Semantic Form
INFL	inflection		SG	singular
INS	instrumental		Slk	Slovak
It	Italian		t	trace
LPT	l-participle		TP	tense phrase
M	masculine		UG	Universal Grammar
			VOC	vocative

Sources

Alx	Alexandreida	VokWeb
AlxB	Alexandreida. Zlomek budějovický, ca. 1350.	VokWeb
AlxM	Alexandreida. Zlomek muzejní, ca. 1350.	VokWeb
Avv.	Avvakum: Žitie, 1672.	RRuDi
BartHla	Bartoloměj Paprocký z Hlahol: Kvalt na pohany, 1595.	DČNK
BawEz	Ezopovy bajky, 1472.	VokWeb
BiblGd	Biblia Gdańska, 1632.	ParBib
BiblBrz	Biblia Brzeska, 1563.	ParBib
BiblKral	Bible kralická, 1613.	ParBib
BiblPozn	Biblia Poznańska, 1991.	ParBib
BiblŠest	Bible šestidílná, 1579-93.	ParBib
BielKron	M. Bielski, Kronika, 1564.	Pepłowski (1966-2004)
BZ	Biblia królowej Zofii, 1455.	PolDi
Chm	Chmielowski: Nowe Ateny albo Akademia wszelkiej sciencyi pełna, 1745.	PolDi
ČNK	Český Národní Korpus	http://ucnk.ff.cuni.cz
DalC	Dalimilova kronika, rukopis Cambridgeský, mid 14 c.	
DalL	Dalimilova kronika, Lobkovický rukopis, 15th c.	
DČNK	Diachronic part of the Czech National Corpus	
Drak	Povesť o Drakule voevode, 2nd h. 15th c.	RRuDi
Dworzanin	Górnicki: Dworzanin polski, 1566.	PolDi
Ezop	Ezopovy bajky, 1472.	VokWeb
FrantPr	Frantovy práva, 1518.	VokWeb
Frol	Povesť o Frole Skobeeve, 1720.	RRuDi
GestaB	Gesta Romanorum: v rkp. Březnickém	
Glatouinus	Simon Eunius Glatouinus: Sepsání kronik a životů, 1565.	DČNK
HavManž	Phaeton (Žalavský) Havel: O ctných manželkách těhotných a rodičkách křesťanských, 1615	DČNK

Hosp	Kryštof Fišer: Knihy hospodářské hospodářství polního, 1705.	DČNK
Hrad.	Desatero kázanie božie: Hradecký rukopis, 1360s.	VokWeb
HvKrJ	Hvězdářství krále Jana, 15th c.	DČNK
IpatLet	Ipat'evskaja letopis', 15th c.	
JanPam	Janczar: Pamiętniki, end 15th c.	PolDi
Kabátník	Martin Kabátník, Cesta z Čech do Jeruzaléma a Egypta, 1491.	DČNK
KodŚwięt	Kodeks Świętosławów (Sulęda), 2nd h. 15th c.	PolDi
KazGn	Kazania Gnieźnieńskie, 1st h. 15th c.	PolDi
KazŚw	Kazania Świętokrzyskie, mid 14th c.	PolDi
Kiev-Pat	Kievo-Pěčerskij Paterik, 1st h. 13th c.	RRuDi
Kit.	Kitowicz: Opis obyczajów i zwyczajów za panowania Augusta III, 2nd h. 18th c.	PolDi
Koř.	Nový Zákon, naps. od Mart. Kořečka, 1425.	VokWeb
Lavr.Let.	Lavrent'evskaja letopis', 1377.	RRuDi
Listy	Listy staropolskie z epoki Wazów, 1601-1665.	PolDi
Lud.	Ludolf's grammar, 17th c.	
Mand.	Cestopis tzv. Mandevilla v překladu Vavřince z Březové, mid 15th c.	VokWeb
Mar.	Codex Marianus, beg. 11th c.	PROIEL
MastDrk	Mastičkář, zlomek Drkolenský, 2nd h. 14th c.	VokWeb
Mer	Merkuriusz Polski, 1661.	PolDi
Milič.	Jan Milíč z Kroměříže: Milíčovský sborník modliteb, 2nd h.14th c.	DČNK
Mill.	Million, český překlad cestopisu Marco Polova, 15th c.	
ModlKunh	Modlitba Kunhutina, beg. 14th c.	VokWeb
ModrzBaz	A. Frycz Modrzewski – C. Bazylik, O poprawie Rzeczypospolitej, 1577.	Pepłowski (1966-2004)
Nikitin	Chożdenie za tri morja Afanasija Nikitina, 1468-1475.	RRuDi
NKJP	Narodowy Korpus Języka Polskiego	http://nkjp.pl
NovBiblKral	Nová kralická bible, 2009.	ParBib
NovgStar	Novgorodskaja pervaja letopis' staršego izvoda, 1260-1330	RRuDi
Op	Opaliński: Satyry albo przestrogi do naprawy rządu i obyczajów w Polszcze należące, na pięć ksiąg rozdzielone, 1650.	PolDi

ParaSoL	Parallel corpus of Slavic and Other Languages	http://www-slavistik. uni-regensburg.de/ parasol
ParBib	Opportunistic Parallel Bible Corpus	
Pass.	Pasionál muzejní, 2nd h. 14th c.	DČNK
PD	Puškinskij dom	http://lib. pushkinskijdom.ru
Podk.	Podkoní a žák, 1409.	VokWeb
PolDi	Polish Diachronic Research Corpus	http://www-slavistik. uni-regensburg.de/poldi
Popr.Rožmb	Popravčí kniha pánův z Rožmberka, 1420-1429.	VokWeb
Posl.	Poslovicy, 17th c.	
PowodPr	H. Powodowski: Propozycyja z wyroków Pisma Św. zebrana, 1595.	Pepłowski (1966-2004)
PřIzJD	Překlad proroků Izaiáše, Jeremiáše, Daniela, end 14th c.	DČNK
PřKT	Beneš z Hořovic: Překlad kroniky Twingerovy, 1445.	VokWeb
PsFl	Psałterz Floriański, 1st h. 15th c.	PolDi
PulKr	Přibík Pulkava z Radenína: Pulkavova Kronika králů českých, ca. 1400.	DČNK
PVL	Povest' vremennyx let (= Nestor's chronicle), redaction unclear	
Rej: Żywot	Rej: Żywot, 1567.	Pepłowski (1966-2004)
Rej: Zwierzyniec	M. Rej: Zwierzyniec, 1562.	Pepłowski (1966-2004)
RozmPrz	Rozmyślanie przemyskie, end 15th c.	PolDi
RRuDi	Regensburg Russian Diachronic Corpus	http://www-slavistik. uni-regensburg.de/rrudi
Skarga	Skarga: Kazania sejmowe, 1597.	PolDi
SkazBoGl	Skazania o Borise i Glebe, Uspenskij sbornik, 12th c.	RRuDi
Sł16	Pepłowski (1966-2004)	
Snář	Hájek Václav z Libočan: Snář, 1581.	DČNK
SNK	Slovenský národný korpus	http://www.snk.sk
Sobieski	Sobieski: Listy do Marysieńki, 1665-83.	PolDi
Sop	Henryk Rzewuski: Pamiątki J. Pana Seweryna Soplicy, cześnika parnawskiego, 1839-41.	PolDi
StaPrze	Staszic: Przestrogi dla Polski, 1790.	PolDi

Starowolski	Starowolski: Pobudka albo rada na zniesienie Tatarów Perekopskich, 1618.	PolDi
ŠtŘb	Tomáš Štítný ze Štítného: Řeči besední, end 14th c.	DČNK
Vesti	Vesti-kuranty, 1656-1665.	RRuDi
Vít	Svatovítský rukopis, 1380-1400.	DČNK
VokWeb	Vokabulář webový	
Wujek	J. Wujek: Nowy Testament, 1593.	Pepłowski (1966-2004)
Zad	Zadonščina, late 14th c.	RRuDi
ŽKrP	Život Krista Pána, 2nd h. 14th c.	DČNK
ŽsvO	Životy svatých otců, 1st h. 15th c.	DČNK

1 Introduction

Slavonic languages are null subject languages, at least in the following sense: Leaving out the subject of a finite verb does not *per se* lead to ungrammaticality, as it usually does in languages like German, English or French.[1] For the sake of illustration, consider the example in (1), from Bulgakov's *Master i Margarita* (MiM), taken from the parallel corpus ParaSoL. While the subject of the embedded adverbial clause is non-omissible, for all intents and purposes, in the English and the German [Ger] version (1-a,b), it is freely left out in Russian [Ru], Polish [Pol] and Czech [Cz] (1-c-e).[2]

(1)   a.   And this the secretary was unable to imagine, though <u>he</u> knew the procurator well.

       b.
Und	das	vermochte	er	sich	nicht	vorzustellen,	obwohl	<u>er</u>
but	that-ACC	could	he	SELF-DAT	not	imagine	although	he

den	Prokurator	gut	kannte. [Ger]
the-ACC.SG	procurator	well	knew

       c.
I	ètogo	sekretar'	prestavit'	sebe	ne	mog,	xotja	<u>Ø</u>	i
and	this-GEN	secretary-NOM	imagine	SELF-DAT	not	can-PT.M.SG	although		also

xorošo	znal	prokuratora.	[Ru]
well	know-PT.M.SG	procurator-ACC.SG	

       d.
Ale	nie	mógł	sobie	tego	wyobrazić,	choć	<u>Ø</u>	nieźle
but	not	can-PT.M.SG	SELF-DAT	that-GEN	imagine	although		not-bad

znał	procuratora.	[Pol]
know-PT.M.SG	procurator-ACC.SG	

       e.
Co	však	následovalo,	tajemníka	ohromilo,	třebaže	<u>Ø</u>	Piláta
what-NOM	however	follow-PT.N.SG	secretary-ACC	bewilder-PT.N.SG	although		P.-ACC

dobře	znal.	[Cz]
well	know-PT.M.SG	

                                                           (ParaSoL)

At the same time, the Slavonic languages are not uniform with respect to the null subject property. A typical pattern is shown in (2), from the same source as (1): While the English and German translations again require an overt pronominal subject in the complement clause (2-a,b), the Slavonic versions omit it (2-d,e) – *except for Ru* (2-c). Ru, at least in its standard variety, differs from Cz and Pol with respect to conditions on null subjects (cf. already Jakobson 1939 (1971; Adamec 1959; cf. also Kosta 1990 for the research history).

---

[1]To be sure, the latter languages also allow for null subjects under (different) restricted circumstances – see chapter 3 and especially section 3.3.2.2 for details.

[2]The referential device in question (pronoun or *pro*) is underlined, here and in the remainder. As usual with parallel corpora, there is no absolute guarantee against "translationese", although these are high-quality literary translations. However, the generalisations which these examples illustrate have been established on the basis of monolingual material (see chapter 3).

(2)   a.   Styopa finally recognized the pier-glass and realized that <u>he</u> was lying on his back in his own bed

   b.   Endlich erkannte er den großen Spiegel und begriff, dass <u>er</u> rücklings
        finally   recognised he the big   mirror   and realised that he backwards
        auf seinem Bett lag [Ger]
        on his   bed lay

   c.   Stepa nakonec uznal   tjurmo   i   ponjal,   čto <u>on</u>
        S.-NOM finally   recognize-PT.M.SG mirror-N.ACC.SG and understand-PT.M.SG that he
        ležit navznič' u sebja   na krovati   [Ru]
        lie-3SG backwards at self-GEN on bed-PRP.SG

   d.   Stiopa rozpoznał   wreszcie tremo   i   zrozumiał,   że Ø
        S.   recognize-PT.M.SG finally   mirror-N.ACC.SG and understand-PT.M.SG that
        leży   na wznak we własnym łóżku   [Pol]
        lie-3SG backwards in own-PRP.SG bed-PRP.SG

   e.   Rozeznal   vysoké   třídílné   zrcadlo   a   došlo
        recognize-PT.M.SG high-ACC.SG three-folded-ACC.SG mirror-ACC.SG and occurr-PT.N.SG
        mu,   že Ø leží   naznak   ve své   posteli   [Cz]
        him-DAT that   lie-3SG backwards in self's-PRP.SG bed-PRP.SG

                                                                              (ParaSoL)

The goal of the present thesis is to show how the variation between Ru, Pol and Cz in this area came about *diachronically*. To this end, we will not only investigate coreferential null subjects (chapter 3), but also alleged subject expletives (chapter 4), and arbitrary (indefinite, generic) null subjects (chapter 5). The latter two phenomena are arguably linked to referential null subjects by implicational relations, both according to typological work and to recent developments in formal syntactic analyses.

According to the facts illustrated in (1)-(2), Ru seems to be situated somewhere between a non-null subject language (like English) and a clear null subject language such as Cz or Pol: On the one hand, a subject pronoun in the embedded clause can be elided, under circumstances to be made precise, as in Cz or Pol, and unlike in English or German. On the other hand, it can also be realised, as in English or German. But note that there is a crucial difference towards Pol or Cz, in the latter case: In Ru, a realised embedded subject pronoun could easily be interpreted as coreferent with the main clause subject (as in (2-c)); in Cz, however, a realised subject pronoun in this position would automatically be interpreted as non-coreferent:

(3)   Rozeznal$_i$ vysoké   třídílné   zrcadlo   a   došlo   mu$_i$,
      recognize-PT.M.SG high-ACC.SG three-folded-ACC.SG mirror-ACC.SG and occurr-PT.N.SG him-DAT
      že on$_{j\neq i}$ leží   naznak   ve své   posteli   [Cz]
      that he-NOM lie-3SG backwards in self's-PRP.SG bed-PRP.SG

> 'He recognized the high three-fold mirror and it occurred to him that he (= someone else) lay on his back in his own bed.'

The status of Ru has been discussed controversely in the literature: Růžička (1986a), Müller (1988; 2006), Kosta (1990) and others view it as a typical null subject language, whereas Lindseth and Franks (1996), Franks (1995), Lindseth (1998), and recently Fehrmann and Junghanns (2008) and McShane (2009) argue for its being a non-null subject language with massive contextual ellipsis. Non-formally oriented approaches tend to give up a ±null subject dichotomy altogether. Stoll and Bickel (2009), who treat Belhare and spoken Ru, based on the theory laid out in Bickel (2003), assume that null subjecthood is a gradual phenomenon, to be subsumed under a continuous measure of *referential density* for texts of a given language. In a Chomskyan tradition, the parametric theory of null subjects has undergone considerable refinements in recent years (cf. the contributions to Biberauer 2010, especially Roberts and Holmberg 2010, for the state of the art): The class of partial null subject languages was introduced, and its relation to subject expletives clarified; research on diachronic change of null subjects in Romance was advanced; and a principal connection between referential null subjects and arbitrary/indefinite impersonals was established. The last point is of special importance for the present thesis, because it theoretically reconstructs the link between referential (chapter 3) and non-referential null subjects (chapter 5).

One of the cornerstones of a parametric approach is the idea that a given parameter setting may derive or explain several otherwise unrelated empirical phenomena at the same time. A number of such correlations with the null subject property have been proposed (see Haider 2001 and Newmeyer 2004 for critique; Roberts and Holmberg 2010). Most analyses, from the most traditional school grammar descriptions to recent theories like Müller's (2006), draw a connection between the *richness/poorness of verbal morphology* and the optionality/obligatoriness of subject pronouns. Ru, Cz, and Pol seem to fit this general idea well: In past tense and conditional mood, Cz and Pol mark grammatical person unambiguously on the auxiliary. Ru, on the other hand, does not distinguish for person on its past tense or conditional forms. As expected, Cz and Pol, but not Ru, are clear, or "canonical" null subject languages.

This partitioning with respect to null subject properties has not been valid throughout the history of the Slavonic languages. The oldest available East Slavonic documents, the Novgorod birchbark charters, show a very large amount of null subjects, which would have to be translated to modern Ru using an overt pronoun. Apparently, all early Slavonic languages were canonical null subject languages, and Ru only later developed into a partial null subject

language. The first main linguistic point of the present work is to show in detail, how and when this development took place, and to what other changes it was related (chapter 3).
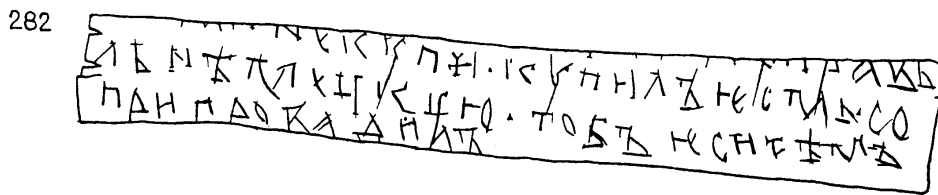
Figure 1.1: Novgorod birchbark charter 282 (1380-1400) (http://gramoty.ru)

(4)   a.   [...] (a so)[li]   ne kupi.   Ø kupil"   jesm'.   so/l'   německuju.
             but salt-GEN not buy-IMP.2SG   buy-LPT.M.SG AUX-1SG salt-ACC German-ACC
             to   Ø b"   jesi   sěm" / priprovadil [Old Ru]
             that-ACC   COND AUX-2SG here   bring-LPT.M.SG

      b.   [...] [A soli]   ne pokupaj: ja   kupil   nemeckuju sol'.
             but salt-GEN not buy-IMP.2SG I-NOM buy-LPT.M.SG German-ACC salt-ACC
             Dostavil   by   ty   ee   sjuda. [Ru]
             bring-LPT.M.SG COND you-NOM she-ACC here

'[...] [But] do not buy [salt]. I bought German salt. If you could bring that.'

(Zaliznjak 2004, 18)

At first sight, diachronic data seem to support a correlation between null subjects and the richness of verbal morphology (cf. Georgieva 1968; Lindseth 1998). The loss of the past tense auxiliaries (and also of the copula in present tense predicative constructions) could have forged the rise of subject pronominalisation (cf. (4-a) vs. (4-b); see already Borkovskij 1949). This is certainly a plausible hypothesis, but we will show that it does not hold uniformly across all grammatical persons. Moreover, the claim is too coarse as it stands, and raises many further questions: Above all, what was the exact time-course of auxiliary drop vs. subject pronominalisation? Was the rise of the present tense null copula a consequence of the loss of null subjects (Lomtev 1956) or a precondition for it (Borkovskij 1978), or were both phenomena a consequence of independent changes in the pronominal system (Georgieva 1968)? What about very early attested examples of auxiliary drop in 3rd person past tense forms? How did the developments relate to register and variety, and how can null subjects co-occurring with an "impoverished" past tense verb form be unproblematic at all in modern colloquial Ru (Weiss 1993)? Was the rise of subject pronominalisation a phenomenon of language contact in 18th c. (O. Yokoyama, according to D. Weiss, p.c.)? Chapter 3 seeks to answer these questions, arguing not only on the basis of isolated examples, but also of quantitative data from corpus samples. The central result is that the explanatory potential

of verbal morphology for the diachrony of Ru null subjects is essentially limited to 1st and 2nd person, whereas further factors have influenced the development in 3rd person. We argue for a *systematic* view on 3rd person null subjects, which takes into account also their non-coreferential (indefinite, generic, arbitrary) readings.

The second main point of chapter 3 concerns the parallel diachronic situation in Pol and Cz. It is only a partially settled issue whether the null subject property was diachronically stable in these languages. According to Pisarkowa (1984, 28ff), early Old Pol sources (*Kazania gnieźnieńskie*, 14th c.) still realised all subjects overtly, and null subjects arose only in the 15th c. We will see in chapter 3 that the text on which this conclusion is based is actually very exceptional. As we show, there is little reason to think that Pol diachronically developed *towards* a null subject language, i.e., in the opposite direction of Ru (which would pose a challenge to many theories of diachronic change).

The null subject status of Old Cz since the earliest texts is less controversial. However, it is a stunning feature of modern Cz that uses optional expletive-like elements, as in (5):

(5) ona   ta   myšlenka   má   něco   do sebe   [Cz]
     EXPL-F.NOM.SG   this thought-F.NOM.SG have-3SG something-ACC   to REFL-GEN
     'This thought has something to it.'   (Lindseth 1994, 336 (10-a))

Across linguistic theories, it has been assumed that canonical null subject languages, which do not even require argumental subjects, show no need for subject expletives either. This claim is implicitly present in functional analyses, such as Lehmann's (1995) – his putative grammaticalization path from demonstratives via subject expletives to verbal affixes suggests that a language will have *either* null subjects *or* expletives. It has also been considered as one of the few valid cross-linguistic generalisations in formally oriented discussions (Haider 2001; but cf. Holmberg 2005 for conflicting evidence). Although Lindseth (1998) shows convincingly that Cz expletive-like elements differ from true expletives by their optionality and their stylistic value, the question remains, how they could arise and develop in a canonical null subject language in the first place. This is our third main point of interest, and the topic of chapter 4.

Trávníček (1961) and, ultimately, Zubatý (1909), have argued that the source of expletive *(v)on(o) 'it'* is not the pronoun *on 'he'*, but rather a deictic interjection *ano* (equivalent to Latin *ecce* 'look!'). This interjection allegedly first became a complementiser and then began to agree with a nominal subject (5). Note that such a development would be quite unexpected under a grammaticalisation view (Lehmann 1995): It would involve a rise (rather than a loss) of inflectional marking, and a move from the (strongly grammaticalised) complementiser to

a quasi-argumental, less grammaticalised element in a lower domain of the clause. Interestingly, Old and Middle Pol also knew an analogous element *ano /(a) ono*, but completely lost it before the 17th c. We discuss the respective changes and propose an alternative view, trying to reconcile the rise of Cz(/Pol) expletive-like elements with their classification as canonical null subject languages .

Chapter 5 of the present thesis deals with *non-coreferential* readings of null subjects in the Slavonic languages. Refined inventories of null elements have been proposed (e.g. by Mel'čuk 1974 for Ru) in order to capture the occurrence of such *indefinite* or *generic* readings. The type of null subjects involved is also often termed *anonymised* or *arbitrary*, and differs semantically from those (coreferential) null subjects, which point to a given, specific antecedent in the context. In chapter 5, we focus on the diachronic development and, mainly, on the systematic relation between referential and non-coreferential null subjects in the three languages.

The interpretations of non-coreferential null subjects overlap only partially in Ru, Pol and Cz – cf. (6) from P. Süskind's *Perfume* (ParaSoL). Among others, Ru 3PL null subjects can support a *generic* reading, as in (6-c), ranging over humans. Cz and Pol usually express this by a reflexive construction or (sometimes) by lexical means, as in the first sentence of (6-e).

(6) a. [How crude! How extraordinarily unsophisticated!] Did one leave diamonds uncut? Did one wear gold in nuggets around one's neck?

  b. Ließ   man Diamanten   ungeschliffen? Trug   man Gold
      leave-PRT.3SG man diamond-ACC.PL uncut   carry-PRT.3SG man gold-ACC.PL
      in Brocken   um   den Hals?   [Ger]
      in nugget-DAT.PL around the   neck-ACC.PL

  c. Razve almazy   ostavljajut neogranennymi? Razve zoloto   nosjat
      PTCL diamond-ACC.PL leave-3PL   uncut-INS.PL   PTCL gold-ACC carry-3PL
      na šee   samorodkami? [Ru]
      on neck-PRP.SG nugget-INS.PL

  d. Czyż nie szlifuje się diamentów? Czy złoto   zawiesza się na
      PTCL not polish-3SG REFL diamond-GEN.PL PTCL gold-ACC hang-3SG REFL on
      szyi w postaci   bryłek?   [Pol]
      neck in form-PRP.SG nugget-GEN.PL

  e. Copak lidé nechají diamanty   nebroušené? Nosí   se   zlato
      PTCL people leave-3PL diamond-ACC.PL unpolished-PL carry-3SG REFL gold-ACC
      ve valounech   na krku?   [Cz]
      in nuggets-PRP.PL on neck-PRP.SG

Non-coreferential 3PL null subjects, in fact, do exist also in Cz and Pol, contrary to the naive expectation that a speaker/hearer of a canonical null subject language would always

interpret such null subjects as coreferential, cf. the newspaper headline

(7) W Mławie trują. [Pol]
     in M.-PRP.SG poison-3PL
     'In Mława, people get poisoned (*lit.* (they) poison (people).)' (NKJP)

But the reading of such examples differs from the one in (6): They refer to an unspecific group of humans *excluding* the speaker. The relevance of genericity and "clusivity" (cf. Filimonova 2005) for the description of the contrast between Cz/Pol and Ru has been noted before in the synchronic literature (Daneš et al. 1987, 237ff; Kosta 1990; Rytel-Kuc 1990; Berger 1991; Weiss 1977); its diachrony, however, is rarely discussed and virtually unknown (*pace* Borkovskij 1978, 220ff). This is where the present work steps in.

Recent work by Holmberg (2010) suggests that the observation illustrated in (6)-(7) may be an instance of a general, cross-linguistic pattern: *Canonical* null subject languages lack generic, speaker-inclusive null subjects (although they may have other non-coreferential readings) in active clauses without special (reflexive, passive) diathesis. *Partial* null subject languages, on the other hand, typically also have null subjects with a generic, speaker-inclusive reading. Holmberg provides a formal analysis which predicts this correlation between the subtypes of null subjects. Interestingly, cross-linguistic evidence for the correlation with diathesis is quite strong (Fassi Fehri 2009).

Assuming that this synchronic analysis is on the right track, we can derive two immediate predictions for diachrony: First, while Ru developed from a canonical to a partial null subject language, the readings of 3PL non-coreferential subjects should also have changed: A generic, speaker-inclusive reading should have been blocked earlier and become possible later. In fact, we should even expect that these two changes proceeded at the same rate (cf. Kroch's 2001 *Constant Rate Hypothesis*). Second, everything else being equal, Pol and Cz should have blocked the generic, speaker-inclusive reading for 3rd person null subjects throughout their history. We will argue in chapter 5 that the first prediction is in fact borne out by our corpus data. The second one, however, at least partially fails: As the evidence collected by Trávníček (1939) for Old Cz and Pisarkowa (1984) for Old Pol shows, 3SG null subjects in active clauses *did* occur in a generic, speaker-inclusive reading in early texts, and were replaced by reflexive constructions only later. But no concomitant changes in the domain of *coreferent* null subjects can be observed. But what does fit the picture is that they soon retrenched in favour of special diathesis marking. This establishes an explanatory, systematic connection between the two phenomena, and provides a good motivation for the otherwise unexplained changes which took place in the realm of reflexive meanings.

The empirical basis of the present thesis is a Ru diachronic corpus (RRuDi) and (smaller) Old Pol (PolDi) one, both specifically designed/compiled and coded for this purpose. For PolDi, we could rely on previously digitised documents from the Academy of Sciences, Kraków, the University of Göttingen, and from an internet library at the University of Gdańsk. For RRuDi, a similar procedure was possible only partially; the rest had to be scanned, treated with Optical Character Recognition (OCR) software, and corrected. It is well-known that within modern Ru, colloquial varieties favour null subjects (and ellipsis in general), to a greater extent than the standard language does (Berger 1989; Kosta 1990; Nichols 1991; Weiss 1993). Attention thus must be devoted to questions of genre and register, even in the diachronic domain. It is instrumental to compare the development in Ru to the diachronically relatively stable situation both in Pol (PolDi) and in Cz. For Cz, the diachronic part of the Czech National Corpus (DČNK) and the corpus of the Academy of Sciences (AV ČR) as extremely valuable sources of evidence could be used. Due to the importance of corpus linguistic aspects for the present work as a whole, we devote a full chapter (chapter 2) to methodological considerations and a presentation of the design, annotation, and evaluation of RRuDi and PolDi.

Although the present work tries to cover the diachrony of Ru, Pol and Cz null subjects in finite clauses in a systematic way, looking at more than only coreferential null subjects, it certainly leaves many blind spots to be explored further. A detailed treatment of *participial* passives and impersonals, which would also have required consideration of the large area of resultatives, was beyond our scope. Some preliminary remarks on the development of *no-* /*to*-constructions in Pol in chapter 5 will have to be sufficient here. The diachrony of true *impersonals* would have been an important related area to look at, but also a large one and, for the most part, complementary to the phenomena discussed here. And finally, the more data could have been considered, the more convincingly could our conclusions have been formulated. Not only are larger digitised text collections called for, but, equally importantly, deeper linguistic annotation of the available ones. It would be our hope that the methodological part of the present thesis established a sensible way for further approaching this task in the future.

2 Methodology and Preliminaries

In Slavonic linguistics, much of the work related to historical and diachronic corpora still has a somewhat pioneering flavour, since ready-to-use, well-balanced reference corpora of Slavonic languages simply do not exist yet (possibly, with the exception of Czech). The purpose of the present chapter is to document the methodological decisions and the preparatory practical work which had to be done before the actual linguistic study could be undertaken. Furthermore, it discusses some important linguistic background definitions which will be taken for granted in the remainder.

2.1 Diachronic Corpus Linguistics

In a sense, historical linguistics is always corpus linguistics: There is no other empirical base for it than the corpus. But of course, not all diachronic linguistics is empirically oriented. Reconstruction, theoretical modelling of language change, typology, and many other fields may do without constant reference to text sources. However, there is a natural connection, which has been developing dynamically in recent years as part of an increasing interest in the empirical base of linguistics.

The evaluation of text collections as favourite pastime of linguists, had been discredited strongly by Chomsky's (1957, 15ff) criticism: (i) Corpora are insufficient in principle, because they can provide only positive, no negative evidence; (ii) Frequency of occurrence is misleading, because it mirrors extralinguistic matters (e.g., that Dayton, Ohio is smaller than New York), rather than linguistic rules; (iii) Linguistic knowledge is independent of frequency of occurrence – even nonsensical sentences which would never appear in a corpus can be judged as belonging or not belonging to the grammar of the language. Interestingly, Chomsky did not deny, in his early works, the value of a corpus as a base for determining an initial grammar, but the grammar should then be able to produce much more than the original corpus, in accordance with the criterion of acceptability for a native speaker (cf. McEnery and Wilson 2002 for a summary; Karlsson 2008 for a detailed documentation of views on corpora in early generative grammar).

There have been various rejoinders to these points from the perspective of modern corpus linguistics. E.g., Pereira (2000) demonstrates that using a suitable bigram model trained on a newspaper corpus, the probability of *Colorless green ideas sleep furiously* will be about 20000 times higher than that of *Furiously sleep ideas green colorless* (Chomsky's 1957 famous example). Stefanowitsch (2006) argues that significant differences in collostructural strength in a corpus – the frequency of a given word in a slot of a given construction in comparison to its alternatives – *do* provide evidence of what is grammatically impossible.

Historical and sociolinguistic corpora were, luckily, exempt from Chomsky's criticism from the start – but only at the cost of diachronic linguistics often becoming ignored by generative grammarians altogether. This has changed since the 1980s (see e.g. Lightfoot 1979; Roberts 1985; Adams 1987; Kroch 1989), and today, formal approaches add to the colourful mixture of theories which try to explain language change. Inevitably, they must also confront themselves with corpus data in this domain. For studies in the presumably dominant diachronic approach today, Grammaticalization theory, corpus data is a more obvious source, although also not compulsory. We will draw from both of the frameworks where appropriate, but mainly focus on their empirical predictions. What remains of Chomsky's criticism as a challenge to grammarians of various theoretical camps, is obvious: The task is not just to document data as such, but to use it as evidence of the changing internal linguistic knowledge of speakers.

Most methods developed for the analysis of modern language corpora carry over to the diachronic domain, but diachronic corpora also come with some specific difficulties and pitfalls. This section discusses the procedures for the construction, coding and annotation, and retrieval and evaluation of the diachronic corpora used for the present work, and some of their methodological foundations.

2.1.1 Corpus Design

2.1.1.1 Text Collection vs. Sampling

Two very different objectives may be distinguished in diachronic corpus linguistics: The *document preservation* approach, as I would like to call it, aims at saving valuable cultural heritage in a digitised form. Large-scale library projects, such as the Monumenta Germaniae Historica (http://www.dmgh.de) or the Manuscriptorium project (http://www.manuscriptorium.com) are scanning their collections and publishing them on the internet; the texts may be read through or – sometimes – queried for single tokens. The second one, the *corpus linguistic* approach is closer to the methods used for modern corpora: Mainly those aspects of the source which may be linguistically relevant are encoded and – if possible – annotated with richer linguistic information. In order to arrive at a balanced and more or less representative corpus, *samples* of the sources may be included rather than full texts. This is the method chosen by the Helsinki corpus of English texts (Kytö 1996) as well as the diachronic part of the Czech National Corpus (Kučera 2002)[1]. While only digitised versions of the most authentic manuscripts really make sense under the "document preservation" approach, reliable editions

---

[1]However, at present the DČNK still offers complete texts rather than samples.

may form useful sources under the corpus linguistic approach. *RRuDi* (*Regensburg Ru*ssian *Di*achronic Corpus) and *PolDi* (*Pol*ish *Di*achronic Research Corpus), which were compiled and coded within the project which with the present work is associated, are intended as being of the corpus linguistic variety (apart from one early menaion), but currently also offer access to the complete source texts.

The two opposing approaches to diachronic corpus building correspond to different goals in the respective corpus-based studies. While researchers in the document preservation tradition often describe one phenomenon in one single text, researchers following the sampling approach usually want to draw conclusions about the language of which the corpus represents a sample, as a whole. Evert (2006) points out that the intended conclusions in the latter case are of an *intensional* kind, i.e., concern something which is not directly (= extensionally) observable, namely, the linguistic competence of human speakers. The hypothesis which drives such a study has to be recast in terms of the existence or frequency of use of an observable phenomenon: How often does the given phenomon occur within a unit of measurement (e.g., a certain number of words, sentences, specific structures)? It is then a matter of linguistic reasoning to decide what this frequency measure tells us about the knowledge of speakers who produced the corpus.

By contrast, classical applications of statistical methods operate on *random samples* drawn from a population. However, it is well-known that most linguistic phenomena are not distributed randomly. Evert (2006, 177) adduces Baayen's (2001) example of the sequence *the the*, which, according to pure chance, should occur roughly once in 300 words (since *the* occurs roughly once within 17 words). But, of course, *the the* can hardly ever be found, except in (rare) cases of error. Randomness is not to be found in language as such, but rather in the choice of a corpus as the source of evidence. This kind of randomness can be illustrated by the *library metaphor* (Evert 2006): Imagine that language, on the extensional view, could be identified with the contents of a gigantic library, which possesses every text ever written (or even uttered) in the given language. A corpus could then be compared to a book or a set of books taken from the shelves of this library. A truly random sample, representative of the library as a whole, could theoretically be achieved by cutting up all the books in the library, putting each unit of measurement – sentence, word etc. – onto one slip of paper, and then draw a certain number of slips of paper from the resulting (thoroughly mingled) heap. A real-life corpus is not usually constructed in this way.

Evert (2006) distinguishes two main sources of non-randomness in a corpus: First, in order to be *representative* of the library as a whole, books for the corpus must be selected in the same proportions as the sections of the library. However, without access to the full library,

we do not know the actual sizes of the sections – say, the amount of spoken vs. written texts, the composition of genres etc. While this first problem may still be partly overcome by a good linguistic understanding of the respective proportions, the second problem is more basic: The unit of sampling (books or parts of books, in the case at hand) is always different from the unit of measurement (sentences, words etc.). As a consequence, phenomena which tend to *cluster* in one text and be absent from another, can create an illusionarily high or low observed frequency if only one of the texts happens to be included in the corpus. Clustering effects are well-known for items of technical terminology (Evert 2006, 184), but they also surface in the main issue of this thesis: While any other Old Pol text from the 14th century represent clear evidence for Pol being a canonical null subject language, the *Kazania Gnieźnieńskie* contain a 3P subject pronoun in virtually every sentence without a full NP subject. At this stage of diachronic Slavonic corpora, we can merely state this as an observation. Evert (2006) demonstrates two general statistical methods of estimating the degree of non-randomness inherent in a given corpus, for a given phenomenon. Although the two problems of non-randomness remain pertinent, the library metaphor is extremely helpful when discussing the design of diachronic corpora as well.

Rissanen (1989; 2008) discusses three problems for diachronic corpus research, the "God's Truth fallacy", "the mystery of vanishing evidence", and "the philologist's dilemma". The first one is a variant of the general problem of drawing a sample: The corpus – even if annotated with all sorts of backgrund information – is always only an inaccurate picture of the language (use) itself. This is especially difficult for historical sources, where the researcher is not exposed to "limitless evidence of the language in everyday life" (Rissanen 2008, 65). The second problem is also well-known e.g. from test design in psycholinguistics: The more variables are to be tested, the larger the group of subjects needs to be. Exactly the same holds of corpora, only this time the size of the whole sample is fixed. Thus, the more factors are used, the smaller and less reliable the evidence will be for a single combination of values. The third problem is related to the fact that a lot of prior knowledge is necessary to interpret the apparent findings form the corpus correctly – "this simple fact is often forgotten because of the attractiveness of and easy access to corpus evidence" (Rissanen 2008, 65). A typical instance of this problem for our object of study would be the relevance of style and variety (OCS vs. Old Ru) across texts for the use of null subjects. Despite these caveats, Rissanen (2008) is, of course, an enthusiastic proponent of diachronic, especially variationist, corpus-based studies.

Of the three problems posed by Rissanen, the first two are directly related to text sampling, and are actually specific variants of problems which belong to the realm of statistic methods

main period	time	no. of tokens (absolute)	no. of tokens (%)
Old English	-850	2 190	0.5
	850-950	92 050	22.3
	950-1050	251 630	60.9
	1050-1150	67 380	16.3
Total:		413 250	100
Middle English	1150-1250	113 010	18.6
	1250-1350	97 480	16.0
	1350-1420	184 230	30.3
	1420-1500	213 850	35.1
Total:		608 570	100
Early Modern English (British)	1500-1570	190 160	34.5
	1570-1640	189 800	34.5
	1640-1710	171 040	31.0
Total:		551 000	100

Table 2.1: Composition of the Diachronic part of the Helsinki Corpus (basic corpus) – Kytö and Rissanen (1993); Kytö (1996)

in general; they can be overcome or at least controlled by the proper use of such methods. What they imply for the design of a diachronic corpus is mainly the demand that a possibly rich mix of genres and a detailed annotation with metadata be included in the corpus.

2.1.1.2 Corpus composition in the Diachronic Helsinki Corpus of English

The Diachronic part of the Helsinki Corpus of English texts (Kytö 1996) has served as a model for diachronic corpora of other languages, partly connected to its further endowment with syntactic annotation (the Penn-Helsinki corpus). The Helsinki Corpus of Scots, the Corpus of Early English Correspondence, the Tycho Brahe Parsed Corpus of Historical Portuguese, and lately the Icelandic Parsed Historical Corpus follow similar corpus plans. The main goal of the Helsinki Diachronic Corpus is to "reflect the principles of socio-historical variation analysis" and to provide "representative coverage of language written of a specific period", including "text types like law, handbooks, science, trials, sermons, diaries, documents, plays, private and official correspondence, etc." (Kytö 1996; Gau 2005, 36f). To this end, the texts are more strongly manipulated than would seem appropriate to many philologists, e.g. "lists of names and longer extracts of foreign language, or verse in a prose text have been omitted" (Kytö 1996). Table 2.1 shows an overview of the corpus composition scheme.

The number of tokens per time "slice" has obviously not been kept constant. Short texts are included as a whole, with extracts of longer texts varying considerably in size; samples

from earlier texts were about 10,000 tokens in size, samples from the late Middle English period onwards between 2,500 and 5,000 words long. Usually two separate extracts were included in a text sample (Kytö and Rissanen 1993, 4). The selection of the texts was primarily guided by criteria like chronological, regional, sociolinguistic, and genre coverage (Kytö and Rissanen 1993, 7). The corpus is based on editions, not manuscripts, and generally avoids translated texts. For datation, the date of the central manuscript of an edition was decisive, not the supposed age of the original; but both are marked up in the annotation. Effort was undertaken to code for dialect and sociolinguistic variables like the age group and social status of the scribe, or the relation between sender and recipient. The text types included are

- *Mundane, non-fictional genres*: Law, Documents, Handbooks/Science, Philosophy, History, Geography, Travelogue, Biography

- *Religious genres*: Homily, Sermon, Rule, Religious treatise, Bible

- *Fictional genres*: Fiction, Romance, Drama, Mystery/Comedy

- *Correspondence* (private/non-private)

Whether this very carefully designed corpus is really representative of the language of its time, cannot be answered, since the size of the target population is impossible to ascertain, and texts are often transmitted or lost by mere historical accident (cf. Claridge 2008; Biber 1993 regarding representativeness in general). But philologically as well as statistically, it seems obvious that a great variation in genres is desirable. In this respect, the Helsinki Corpus of English Texts can definitely serve as a model. The issue of varying sizes of subcorpora, however, may present a problem for the application of certain quantitative methods.

2.1.1.3 Corpus composition of the Regensburg Russian Diachronic Corpus (RRuDi)

The dependence of results in corpus linguistics on the sampling of the corpus is so obvious and well-known that we need to provide a proper justification for the selection of texts in our corpus. To this end, standard literary histories of (Old) Russian, works on Old Russian linguistic textology, and anthologies of Old Russian literature were tapped as sources of knowledge, which (to a certain extent) represent common practice in the field. As a result, we arrive at a corpus plan which is founded in the results of linguistic and literary research, rather than in accidental external factors.

On Schmidt's (2002) view, what is usually referred to as "Old Russian literature", should rather be viewed as the common literature of the East Slavonic "Slavia orthodoxa" until the

end of the 16th century, i.e. as a common heritage of (later) Belorussian, Ukrainian, and Russian. Medieval literature (from Vladimir's baptism in 988 until the second half of the 17th century) in this area was distinguished by its strong adherence into theological and clerical functions; classical literature remained almost unknown for a long time. Texts from this domain, as e.g. liturgical literature, readings from the gospel, hymns, canon law etc. form the largest corpus, followed by the writings of (orthodox) church fathers such as Ioann Chrysostomos. The amount of clerical literature, which was generally translated from Greek into a South Slavonic variety and copied throughout the "Slavia orthodoxa", may be estimated to 90% of the preserved manuscripts (Schmidt 2002, 2). Marti (1989, 388) finds that more than 75% of all preserved manuscripts exclusively consist of clerical texts. Schmidt (2002, 6ff) arranges the early *clerical* literature into the following genres:

- oldest preserved East Slavonic manuscripts from the 11th century : *Ostromirovo evangelie* (1056), *Izbornik Svjatoslava* (1073), *Izbornik* (1076)

- hagiographic texts (*žitie*): (i) lives of saints: *Žitie Alekseja čeloveka božija, Skazanie i strasť i pochvala svjatuju mučeniku Borisa i Gleba* and *Žitie Feodosija Pečerskogo* (both part of the *Uspenskij sbornik*, 11th/12th c.) etc. – (ii) lives of (partly canonised) heroes and rulers: *Povesť o žitii Aleksandra Nevskogo* (13h c.) etc. – (iii) collections of dicta and narratives about monks (*pateriki*): *Paterik Kievsko-Pečerskogo monastyrja* (1st h. 13th c.) etc.

- homiletic literature (sermons, *slova*): e.g. by *Feodosij Pečerskij* (11th c.), *Grigorij Belgorodskij* (12th c.), *Kirill Turovskij* (12th c.)

- hymns: byzantine liturgical chants of rich poetic and complicated formal structure

- apocryphal texts: partly canonically allowed (*istinnye knigi*), partly forbidden (*ložnye, otrečennye knigi*) legends, collected in menaia such as the Great Reading Menaia of metropolite Makarij (16th c.); e.g. *Choždenie bogorodicy po mukam*

- clerical letters (epistolae, *poslanija*)

- pilgrimage reports (*choženija*): *Choženie igumena Daniila* (beginning 12h c.)

The *mundane* literature between 12th and 14th c. comprises

- historiography: *Novgorodskaja pervaja letopis' staršego izvoda* (Novgorod primary chronicle) (13th c.), *Povest' vremennych let* (Nestor's chronicle) (beginning 12th c.), regional chronicles, *Iudejskaja vojna* (translated from Josephus Flavius) (12th-13th c.) etc.

- phantastic, entertaining narratives: *Skazanie ob Indijskom carstve, Povest' ob Akire*

- epos: *Zadonščina* (end 14th c.)

- isolated, unclear or mixed genres: *Slovo i molenie Daniila Zatočnika*

According to Rothe's (2000) pointed summary, the only really "Old Russian" *literary* work of notable size here is the chronicle. Everything else is more or less influenced by OCS (or not literary). This is somewhat in line with earlier somewhat polemic accounts like e.g. Issatschenko (1980), who bases his overview of the Russian textual tradition on the concept of a general diglossia: Gospels, psalterions, liturgical literature, sermons, legends of saints, as well as theological and religious, but also entertaining literature were all written in (Old) Church Slavonic. Juridical and administrative charters, treaties, donations, testaments, laws, etc. were composed in the East Slavonic vernacular (Issatschenko 1980, 68). He singles out the following main threads:

1. Kiev period and transition to Moscow period (11th-14th c.): The high literature was oriented towards Byzantine literary style – *Izbornik Svjatoslava* 1073 and 1076, *menaia*, works by *Kiril Turovskij*, the *Ostromir* and other gospels. Juridical texts *(Russkaja Pravda*, several charters, treaties etc.) from the north (Novgorod, Pskov, Smolensk) serve practical purposes and are written in a formulaic vernacular, characterized, e.g., by paratactic syntax and referentially unclear pronouns. Early private correspondence and notes on birchbark (Novgorod, Pskov) (XI.-XIV.) display the vernacular used by common citizens, sometimes permeated by Church Slavonic elements. Issatschenko thinks their value for Russian linguistic history rather low (p. 67), quite the opposite of what is the commonly accepted view today (Zaliznjak 2004). Chronicles (since 11th c.) – usually classified as clearly East Slavonic – rather tend towards Church Slavonic, on Issatschenko's view. The *Slovo i molenie Daniila zatočnika* (13th c.), basically Church Slavonic, is permeated by vernacular elements for humorous effects.

2. Moscow period (14th-15th c.): Hagiographic and also some historiographic works are marked by the artificially enforced rebulgarisation of the literary language. Charters and correspondences in 15th-16th c. are written in an East Slavonic variety, secular prose varies in register (partly Church Slavonic: the *Tale of Dracula*, mostly East Slavonic: the *Choždenie za tri morja* by Afanasij Nikitin).

3. 16th c.: Early grammar books; correspondence by Maksim Grek and Ivan the Terrible varied in register, but mostly extremely elevated; Moscow chancellery language; first conversation books by foreign visitors to Russia.

4. 17th c.: Many sources display a strong Polish and/or Ukrainian influence; Ludolf's grammar; secular prose and some drastic satires; Avvakum's writings are not considered as representative of natural and unartificial, folk language by Issatschenko (p. 330ff), but rather as a mixture of styles; the *Vesti-kuranty*, early newspaper-like notes of small circulation, translated form Polish or West European languages.

Taking overviews like these into account, a first selection of relevant texts was devised for the RRuDi corpus by Ernst Hansack within a research project at the University of Regensburg. The following genres were singled out:

- Chronicles: Chronological history of the Russian state and its rulers, with entries for every year. Linguistically usually viewed as (Old) Ru with some Church Slavonic characteristics.

- Charters and juridical texts: Treaties, testaments, law, jurisdiction. Linguistically (Old) Ru, almost no Church Slavonic features; often formulaic, with little variation.

- Correspondence: Letters, notes, messages, semi-official "public" letters. Linguistically reaching from the purely East Slavonic Novgorod birchbark charters up to the stylistically refined, "high variety" letters by Ivan the Terrible and Lord Kurbskij.

- Travel literature: Narratives by Ru merchants, adventurers and messengers who report their experiences from expeditions to foreign countries, often commenting on unusual ways of living, religion, and politics. Linguistically (Old) Ru, with different amounts of "high variety" / Church Slavonic features. Since 15th c.

- Homilies: Religious instruction, often in poetic style; many Church Slavonic characteristics. Mainly 11th/12th c.

- Counsellors' literature: Religious and mundane advice on leading a proper Christian life; (Old) Ru with a medium amount of Church Slavonic characteristics.

- Vitae: Life stories of saints, often intended as role models in religious education. Strong Church Slavonic influence. Since 11th c.

- Fictional prose: Fictional or partly fictional tales, like *Dracula*, *Frol Skobeev* etc.

- Translations: Usually viewed as sort of a genre, but actually containing historiographic texts as well as homilies and other purely religious texts.

This extended research corpus now consists of the texts given in table 2.2.

text	origin (c.)	genre
Vita of Ioann Chrysostomos	10th	biographical, religious
January Menaion	10th	religious, sermons
Flavius Josephus: Iudejskaja vojna	11th	translation, historical narrative
Šestodnev Ioanna Èkzarcha (Ed. Barankova)	11th	religious, instructional, non-fictional prose
Nestorʻs chronicle (Lavr.)	11th	chronicle
Izbornik Svjatoslava (1073), excerpts	11th	legends, religious advice
Izbornik Svjatoslava (1076), excerpts	11th	religious, mixed, non-fictional prose
Pandects of Antiochos	11th	sermons
Novgorod birchbark charters, excerpts	11th-15th	letters, notes, treaties etc.
Novgorodskaja kormčaja / Pravda russkaja	11th	juridical
First Novgorod Chronicle	11th	chronicle
Poučenie Vladimira Monomacha	12th	mundane advice
Slova i pritči Kirilla Turovskogo	12th	sermons
Slovo i molenie Daniila Zatočnika	13th	sermons
Kievskij Paterikon	13th	biographical, narrations about monks
Sinaijskij Paterik	13th	religious, biographical
Tale of the Siege of Pskov	14th	historical
False Dimitrij by Avraam Palicyn	14th	historical, partly fictional prose
Žitie Alexandra Nevskogo	14th	historical, biographical
Court charters	14th-17th	protocols, treaties, testaments etc.
Choždenie za tri morja A. Nikitina	15th	travel report
Sermons of Serapion from Vladimir	15th	sermons
Domostroj by Pope Silvester	16th	advice, mundane and religious
Vita of Sergius from Rodonež	16th	biographical, religious
Zadonščina	16th	epos
Correspondence of Ivan the Terrible	16th	letters in elevated style
Tale of Dracula	16th	mundane legend, partly fictional prose
Žitie by the Protopope Avvakum	17th	biographical, mixed
Vesti-kuranty	17th	predecessor of newspapers
Tale of Frol Skobeev	18th	fictional prose

Table 2.2: Corpus plan of RRuDi

2.1.1.4 Other available Russian historical corpora

The most technically advanced corpus of early Old Ru texts is presently offered by the project Manuskript (http://manuscripts.ru) at the University of Iževsk, headed by V. Baranov. The corpus is focused on the early period (until 14th c.), but also contains writings of *Lomonosov* and is diverse in genres. Access is free for research purposes, and the texts are increasingly also lemmatised, and beginning to be annotated with morphosyntactic information. The quality of the corpus and of its online presentation is tremendous, the only drawback which may be found at all concerns a certain technical closedness, the lack of adherence to international encoding standards, including Unicode, which sometimes hinders cooperation.

The Novgorod birchbark documents have been made available for reading (but not querying) online, and offer a convenient way of exploring these texts, including a tokenisation and translation into modern Ru, for many charters. The website http://gramoty.ru, run by the Academy of Sciences Institute of Russian language, is making available chronicles and other larger Old Ru texts, first as scanned pages of editions, now increasingly as tokenised and searchable texts. A morphosyntactic tagging is under way.

Not strictly a corpus, but an online readable collection of important texts, is the corpus of the Academy of Sciences Institute of Russian Literature (*Puškinskij Dom*) http://lib.pushkinskijdom.ru is addressed at a broad public and presents its contents in aligned parallel paragraphs of Old Ru and a Modern Ru translation. Since the texts are partly patched up from various sources and editions, they have to be used with care. Editorial comments are provided for all of the texts, including the manuscript sources and principles of "correction" (*ispravlenie*) or adaptation. For some purposes the aligned translations are extremely valuable, and some texts have not been edited in any other collection (such as e.g. the *Devgeniovo dejanie*). Therefore, we included those texts which were transcribed from a single, reliable manuscript, possibly with minor additions from another, in an oportunistic parallel corpus, but ignored texts which obviously stem from mixed sources or underwent undocumented "corrections" by the editor. Given that Zaliznjak's (2008) corpus contains several of these texts, we think that this procedure is justifiable.

RRuDi sets itself apart from these strong and fruitful initiatives by insisting on encoding standards and corpus-linguistic methods of analysis and annotation, by its broad diachronic coverage and its readiness to work with text samples and quantitative methods, rather than preserving rare and valuable documents.

2.1.1.5 Polish Diachronic Research Corpus (PolDi)

The second corpus for which the project with which the present work is associated carries responsibility, is the Polish Diachronic Research Corpus PolDi. PolDi consists of the Extensible Markup Language (XML)-annotated texts of the *Korpus staropolski* of the Academy of Sciences in Kraków, which contains virtually all relevant texts before 1500; the so-called Göttingen corpus of Polish Baroque, digitised editions of middle Pol texts, which were kindly mady available by G. Hentschel, Oldenburg; and some digitised texts from online virtual libraries. Its composition roughly follows the choices made in standard anthologies of Pol language history and may be called opportunistic. Nevertheless, the corpus is the largest diachronic resource available as a corpus for Pol at the moment. Since Pol libraries are massively digitising their collections, it is foreseeable that a truly enormous corpus will be viable in the near future, provided that Pol computational linguists will be willing to work on OCR and linguistic annotation of these sources.

2.2 Corpus Implementation, Coding and Retrieval

2.2.1 Data format

Edited textual data can be structured in many unpredictable ways: It may contain comments, additions and corrections at various levels, e.g. regarding sentences, word forms, or graphemes; there may be various possible regularizations for some original forms, but none for others; some, but not all of the original data may be aligned with parallel texts in other languages, sentence by sentence or by larger paragraph units. Data of this kind is called *semi-structured*. The most widespread format for such data is the so-called Extensible Markup Language (XML), an international standard primarily intended as a storage and exchange format, but – with the increasing power of databases – also more and more used as a data format proper. Due to the linguistic and textological specifics of diachronic corpora (cf. Meyer 2005 for details), the data format of choice for the representation of RRuDi was a *multi-level standoff* XML format. Such a format can best capture the fact that the annotations for various pieces of linguistic and non-linguistic information may overlap in scope. In a multi-level standoff XML format, the levels may in principle be manipulated independently from each other, leaving the rest (and the base text) intact. The mature XML format PAULA (http://www.sfb632.uni-potsdam.de/~d1/paula/doc/), viz. Dipper (2005); Wörner et al. (2006) has all the desired properties, plus the additional advantages that its relation to the XML formats of various annotation tools have been explicitly documented, that the respective

conversion tools are available, and that it serves as (primary) input format for the powerful corpus server/web interface Annis-2 (Zeldes et al. 2009).

Each of the freely definable, independent levels in PAULA contains annotations for (not necessarily continous) segments of primary data, the so-called *markables*. The first task, therefore, is to specify a sensibly restricted set of annotation levels, each of which represents a well-formed XML structure, i.e. a hierarchical structure without crossing edges. The problem of crossing edges is immediately apparent if we take into account the necessity in diachronic linguistics to properly cite example material including its location in the source text: Physical units such as page breaks overlap with text structural units such as paragraph breaks or verse breaks, as well as with linguistic units such as sentence breaks (to the extent that these can be recognized).

The following annotation levels form a minimal requirement fo diachronic corpus texts:

1. Physical structure of the document

2. Textual structure of the document

3. Token annotation

4. Higher-level linguistic annotation

In corpora of *modern* languages, 1.-2. are often restricted or left out completely, while mainly 3.-4. are elaborated. For historical corpora, 1.-2. cannot be dispensed with. Much more clearly than in modern corpora, the *tokens*, i.e., the minimal units[2] for linguistic annotation in our corpus, already belong to a level of abstraction. Tokens may be split physically across lines or pages, marked by a hyphen or not. The actual orthographic shape of a token may vary drastically, involving abbreviations, sub- and superscripts. The intended tokenization of a manuscript may also not be obvious, due to the lack of spaces between words (*scriptum continuum*), leading to an ambiguous stream of tokens. Ribarov et al. (2004), in their "Annotated Corpora of Text" (ACT) XML format, propose a level of "original forms" (*oforms*, tokens at the original orthographic level) and "regularized forms" (*rforms*, basic units for linguistic annotation) in order to deal with such difficulties. The relation between *oforms* and *rforms* is a one-to-many mapping; thus, a span of orthographic text may be split into rforms in several ways, allowing for ambiguous (or even partial) tokenization. The label of an *rform* is, in principle, independent of the original orthography, but in the ACT annotation tool, a cascade of regular expression replacements can be devised to derive it from the latter.

---

[2]There has been discussion on implementing a subtoken level in Annis, which would be needed e.g. for morpheme markup. Nevertheless, the token remains a designated level to which most of the linguistic analysis is anchored.

It is an axiom of the Paula annotation scheme in its present shape that the token level is unique. In the Annis query language, searches which leave their annotation level unspecified refer to the token level by default. Other annotation levels, however, may be mapped to the token level freely, e.g. to express ambiguous assignments of regularized base forms (*rforms*). For the RRuDi and PolDi corpora, we decided to adapt the rationale of the ACT setup to the Paula annotation scheme in the following way: The designated token level of Paula and Annis corresponds to the *oform* level in ACT. I.e., tokens are assigned directly to continous spans of the original orthography. In the case of editions, no ambiguities arise, unless the editor notes possible alternatives to his own tokenization. The level of *rforms* will contain regularized forms for all the orthographic tokens, constructed semi-automatically from the original orthography, convenient for querying. In manuscripts with *scriptum continuum*, spans which may be split into word forms ambiguously, are left as one connected piece (*oform* in ACT, *token* in the sense of Paula) and further levels of *rforms* are invoked to contain the alternative tokenizations. A potential problem is that it cannot be known in advance at which *rform* level a certain tokenization will be located; therefore, the interface must run a query for *rforms* on *all* respective levels (*rform1, rform2, ..., rformn*) simultaneously. Alternatively, all the *possible* tokenizations might be collected into an abstract level of markables, say, *rform* (in general) which would then be searched in one swoop. Naturally, ambiguous tokenization will often give rise to ambiguity at higher annotation levels as well. Such ambiguities may, again, be easily represented by separate levels of annotation over markables, which would be queried sequentially by the interface, and could also be collected into an overall annotation set.

2.2.2 Annotation tool

The ideal annotation procedure envisaged for RRuDi (and PolDi) was supposed to be *semi-automatic*: As far as possible, automatic methods should provide candidate markup, which was then corrected manually. In the optimal case, this should turn into a very flexible process which would allows for additional annotation/correction loops if necessary.

As far as manual annnotation is concerned, important requirements include the possiblity to restrict annotations to a set of pre-defined categories, which may ideally be selected from menues or via buttons and the like, rather than being typed in; furthermore, views of the annotated data as running text and in a tabular "score" form have proven helpful for the annotator (cf. Dipper et al. 2004). The most widespread tools for linguistic annotation lack at least one of these parameters: EXMARaLDA (http://www.exmaralda.org/) is based on the score presentation format, but offers no pre-definition and menue-selection of anno-

tations; MMAX (http://mmax2.sourceforge.net/), on the other hand, lacks the score format, but allows for the latter (cf. also Burghardt 2008 for a comparison of many current tools for linguistic annotation).

The tool coming closest to Dipper's et al. requirements is probably ELAN (http://www. lat-mpi.eu/tools/tools/elan), which – in its latest versions – provides both running-text and score format, and even allows for a lexicon of possible annotations per annotation tier. Note, however, that neither of these programmes offers an easy way to integrate external software, such as normalisers, tokenisers or taggers. The stepwise procedure alluded to above would thus mean repeated export into suitable formats and re-import after external processing, which requires programming the respective converters and brings about a considerable risk of errors and data loss.

Since a guided annotation was deemed most important for the current task, first steps at building RRuDi were undertaken with MMAX, including quite a lot of programming around this core: converters from the Microsoft Word XML format WordML[3] into MMAX stand-off XML, as well as from other, previously used XML formats into MMAX XML; interfaces between the MMAX XML format and external tools such as stochastic taggers and morphological analysers. After a short period of time, however, it became clear that MMAX was not prepared to handle the sometimes very large documents, which could not easily be split without losing annotation information. Furthermore, the MMAX strategy of applying formatting sheets (so-called XSL sheets) in order to make annotations visible, was extremely time-consuming. For these reasons, the annotation tool was switched.

A tool which unites all the above features, is very stable and powerful and, additionally, comes with numerous language processing components of its own, is GATE (http://gate.ac. uk/). This is the tool on which the project finally settled. The only remaining problem, then, was the link between GATE's XML format and PAULA. This bottleneck could be overcome by (i) reducing GATE annotations the relevant subset with the help of so-called JAPE rules[4], (ii) exporting into external XML using a plugin devised by the programmers of the American National Corpus, and (iii) converting externally into the structurally closest format which has an importer to PAULA, namely, EXMARaLDA XML, via a python programme written for this purpose.

---

[3] The results of scanning and OCR were Microsoft Word files.

[4] JAPE is a rule format native to GATE, which incorporates not only regular expressions, but even fully-fledged Java programming for annotations.

2.2.3 Annotation strategy

Semi-automatic annotation of the kind envisaged for RRuDi could mean that, e.g., a preliminary morphosyntactic tagging would lead to some useful tags, but also to systematic errors, which could then be filtered out partially with the help of rules over annotations, and partially by manual correction of single instances. Or one specific feature value might be automatically added by default: E.g., a pro-drop subject could be ascribed automatically in some texts unless any referential subject or other null subject has been previously annotated in the clause. GATE's flexibility was enormously helpful here.

Since texts in RRuDi and PolDi stem from various sources and formats, and since the intended annotation was complicated by itself, corpus encoding was a far from trivial task and took up much more time than is usually the case with modern corpora. Manuscripts had to be typed in, usually using a text processor like MS Word. Scanned editions had to undergo OCR and manual correction, leading to a MS Word or OpenOffice document which included remarks, footnotes, line numbers, page marks etc. within the text proper. These were then marked up within MS Word using a set of pre-defined character and paragraph styles, and imported from the respective Word XML into GATE. Within GATE, the tokenised texts first had to be normalised orthographically, using a python programme which was linked up with GATE via the TaggerFramework plugin. A large part of the normalisation was programmed by Christine Grillborzer. Some of the material, e.g., the *Domostroj* edition digitised by Mirjam Zumstein, already came with manually annotated normalisations, which were imported as annotation types of their own (*rform*) into GATE. In single cases, permission was gained to use electronic internet collections (e.g., http://izbornyk.org.ua/), which required the parsing of html pages into a suitable XML format for import into GATE.

Annotation with linguistic categories is essential for most corpus-based research into syntactic phenomena. Possibilities abound, reaching from part-of-speech (POS) tagging down into morphology and phonology and up as far as to rhetorical relations. Quite obviously, the more can be automatised (with decent quality), the deeper and more detailed the annotation can become. But how can we hope to automatise the annotation of a historic corpus, given the problems discussed above and the general lack of processing tools in this area? The method presented in this section is based on two major resources: First, automatic morphological analysers for the respective languages, which may also be used as form guessers for words which are not included in the (necessarily small) lexicon. Second, information drawn from alignments with modern translations at the word level. Morphological analysers can be programmed using finite-state frameworks such as the Xerox FST tools (Beesley and Karttunen 2003) or INTEX/NOOJ (Silberztein 2000; 2005) in a manageable amount of time. The

potential of word alignment for the annotation of historical texts cannot be overestimated, as long as sufficiently large parallel corpora with original texts and their modern translations are available. In the case of Russian, the huge parallel text collection at the Academy of Sciences' Institute of Russian Literature (*Puškinskij dom*) is a good starting point.

For untokenised RRuDi (and PolDi) texts, a first tokenisation (automatic split into word forms) was done using the ANNIE tokeniser and sentence splitter tools coming as part of GATE. Then, automatic part-of-speech tagging was conducted, in one of three variants:

1. using a cascade of automatic finite-state morphological analysers for Old Church Slavonic (OCS) and Old Ru, devised especially for this purpose, and an external one for Modern Ru

2. (for translated texts:) using the finite-state morphological analyser for Old Ru in combination with the tag information on aligned Modern Ru texts

3. (for Polish texts:) using a Modern Pol tagger (Morfeusz, cf. http://sgjp.pl/morfeusz/)

The result was used as input for a whole battery of flexible rules, in order to facilitate the annotation of categories relevant for the present study. E.g., it was helpful to be able to recognize most finite verb forms properly, where corrections were only rarely needed. The rest of the annotation had to be done mostly by hand, but again, the possiblity to flexibly define rules and little programmes over annotations greatly helped in the process.

2.2.3.1 Finite-state morphology for historical languages

Both the OCS and the Old Ru morphological analyser was written using the Xerox finite-state tools (Beesley and Karttunen 2003). OCS morphology is rather well-described and uncontroversial. A convenient and succinct source was Diels (1963). The Old Ru inflection module implements Old Ru morphology as presented in Franz and Tuschinsky (1982); Schmalstieg (1995); Borkovskij and Kuznecov (1965). Each programming tasks required extensive testing and correction, and took about a month to complete. The morphology modules are completely independent of the annotation task at hand, and can be built into other software if needed. E.g., the OCS module has been linked with a web application which analyses forms or generates full paradigms and can be used for learning purposes and is available under http://rhssl1.uni-regensburg.de:8080/OCS – see figure 2.1.

The OCS analyser has the great advantage of being integrated with a relatively substantial digital lexicon, for the most part prepared by Christine Grillborzer from a scan of Sadnik-Aitzetmüller and Aitzetmüller (1955). The Old Ru morphology has its weaknesses in the lack of sufficient lexical resources, but it can still always be used as a rule-based guesser.

Figure 2.1: OCS finite-state morphology embedded in a web application

2.2.3.2 Alignment of historical/modern parallel texts

In a translated historical corpus, we face at least one alignment relation, namely between the segments of the historical source and their counterparts in the modern version. Depending on the kind of historical source, there is often another, implicit alignment hidden behind the scene: that between the actual manuscript or print and its edited, possibly orthographically normalised, version. This state of affairs holds e.g. for the extremely rich edition of Old Ru texts published in the *Biblioteka drevnej Rusi* (Lichačev et al. 2000), which has been made available online by the *Puškinskij dom* (http://lib.pushkinskijdom.ru). Assuming that we can automatically produce reliable and sufficiently fine-grained alignments between these three versions, we may collect information on orthographic variants and their normalised counterparts and on the relation between (normalised) Old Russian and (tagged) Modern Russian. The situation is comparable to one familiar from endangered languages: A resource-poor language can profit from the information tacitly implied in translations into a resource-rich language. My strategy for tagging Old Ru in this vein was to first run the finite-state guesser for Old Ru, resulting in a (sometimes very large) number of theoretically possible tags for a given form. Since the lexicon is small and the root has to be guessed, proposals vary greatly in quality. As a next step, the information available on the aligned word in the Modern Ru

45

version is exploited. It is instrumental to carefully design transfer rules, which guide the disambiguation in a sensible language-specific manner. The details of this whole procedure are laid out in Meyer (2011).

2.2.3.3 Using a modern tagger on PolDi

For PolDi, a different approach to annotation was chosen. Morphosyntactic tagging was conducted with the modern Polish tagger Morfeusz (http://sgjp.pl/morfeusz/). The GATE annotation framework provides a so-called GenericTagger plugin which could be adapted to use Morfeusz via a python programme. A major obstacle (actually, rather a feature) was the fact that Morfeusz uses linguistically, rather than orthographically, sound principles of tokenisation. I.e., an orthographic form like *myślałbym 'think-*COND.1SG*'* would be assigned the token triple <*myślał* (=PRT.SG.M of *myśleć 'to think'*), *by* (COND), *m* (= clitic 1SG of *być 'to be'*)>. Even worse, under this approach, a form like *piekłem* becomes *ambiguous* between two tokenisations, namely <*piekłem* (=INS.SG of *piekło 'hell'*)> and <*piekł* (=PRT.SG.M of *piec 'to bake'*), *em* (= clitic 1SG of *być 'to be'*)>. Thus, the linguistic facts call for ambiguity in tags, but also in tokenisations. As a result, each solution by Morfeusz had to be transferred to one sequence of possible linguistic tokens (per orthographic token), marked up for lemma and morphosyntactic categories, as illustrated in figure 2.2.

2.2.3.4 Further annotation

Having the historical texts lemmatised and tagged, at least for those tokens which are equal or similar to their modern counterparts, can be valuable for a number of purposes. In our case, the real trade-off from this procedure was the possibility to formulate rules for the automatic pre-annotation of the categories which interested us, namely, potential null subjects. Thus, a rule for PolDi was devised to the effect that every sentence which contained no MorfeuszToken with a nominative case tag, got marked as a potential null subject clause, with an agreement marking corresponding to the features of its finite verb. The respective finite verb was marked as having a pro-drop highest argument. In the manual correction phase, these markups could be controlled more quickly. Cf. the listing in Algorithm 1 for an illustration of such rules, using the Java Annotation Patterns Engine (JAPE) format impemented in GATE (http://gate.ac.uk/).
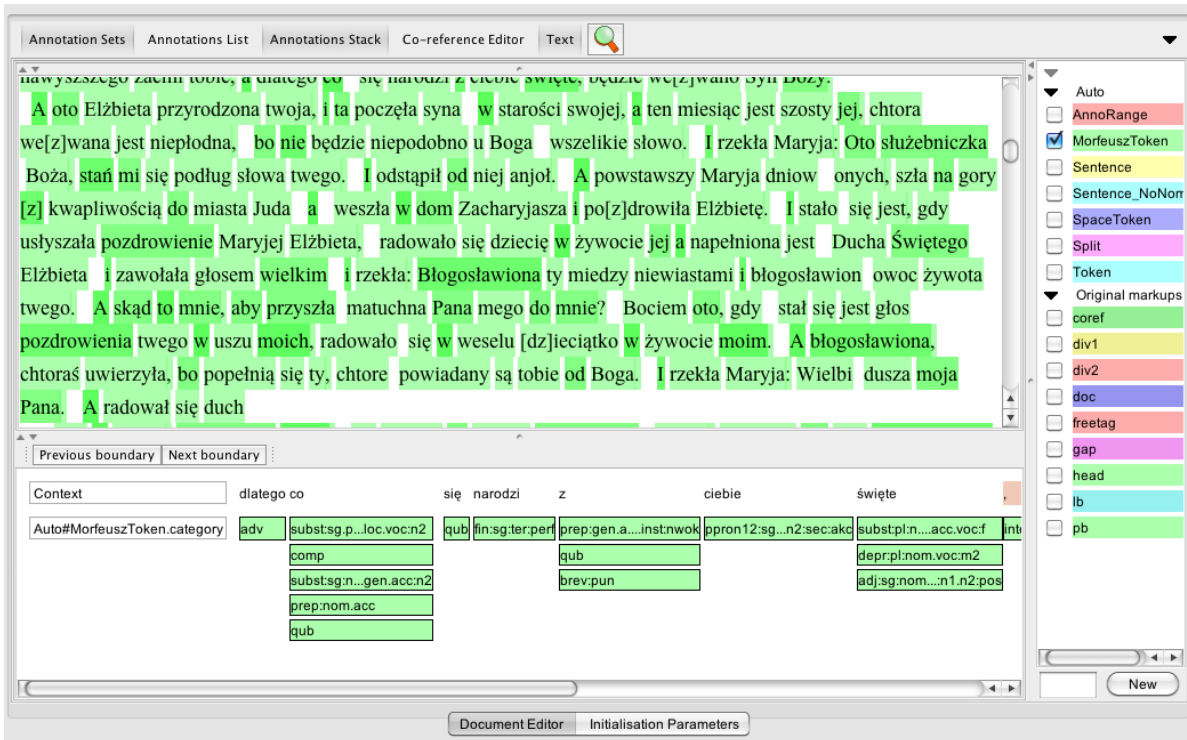
Figure 2.2: Ambiguous tags output by Morfeusz

Algorithm 1 JAPE rules for automatic markup before manual correction

```
Phase: select0
Input: MorfeuszToken coref
Options: control = all
Rule: MT_1
({MorfeuszToken.category    =~    "^(fin\\:pl\\:ter|praet\\:pl)",
!coref}):morf0
--> :morf0.coref = {ref_type = "pro-definite-3pl"}


Phase: select1
Input: MorfeuszToken coref
Options: control = all
Rule: MT_2
(({MorfeuszToken.category =~ "^praet\\:sg", !coref}):morf1
({MorfeuszToken.lemma == "by"})?
{MorfeuszToken.category =~ "^aglt\\:(sg|pl)\\:(pri|sec)"})
--> :morf1.coref = {ref_type = "pro-definite-1/2"}


Phase: select2
Input: MorfeuszToken coref
Options: control = all
Rule: MT_3
({MorfeuszToken.category    =~    "^(fin\\:sg\\:ter|praet\\:sg)",
!coref}):morf2
--> :morf2.coref = {ref_type = "pro-definite-3sg"}
```

Furthermore, instances of pro-drop from the non-corrected parts of the texts, could be found, at least with a feasible amount of erroneous hits. It is obvious that most data categorisations necessary for the present work had to be taken by a human annotator. In the ideal scenario, however, this annotator would not just proceed word-by-word, but s/he could work through a set of automatically provided candidate annotations systematically. E.g., when determining the reading of the reflexive marker *się*, it speeds up the annotation process enormously if one can just jump from instance to instance, clicking off the respective categorisations. It is exactly this type of strategy that GATE supports.

2.2.4 Evaluation via JAPE and R

One of the many advantages of GATE is the fact that it is flexible enough to be utilised as a local corpus manager. Thus, text and annotations could be exported into tables suitable for the statistics software R (http://www.r-project.org/), in order to run descriptive statistics, tests, and plots for the purposes of this thesis. In order to arrive at a data frame suitable as input for the statistics software R, the coded corpus was run through a JAPE script which extracted, from each document,

- basic metadata: document name, time of origin (lower and upper bound), genre code

- for each clause within a randomly selected token range of fixed size, the features and values assigned for

    - coreferential properties of the subject

    - morphosyntactic marking of the finite verb

    - grammatical properties of a passive participle or reflexive verb form (if such existed)

- the text of the clause itself, for later reference

and outputs this information in a tabular format, to be read back into R.

Eventually, import of the corpus into ANNIS-2 and querying via the web interface with its statistical capacities and exporters into statistics software, allows for the same analyses. A clear advantage of ANNIS would be the mature query language which is much better suited for linguistic needs. E.g., it was quite difficult to express in JAPE that when extracting coreference markup from a given clause, I really only needed verbs and subjects from this clause as a data point, rather than categories from clauses *embedded* in it. In ANNIS, this is very easy to define.

Nevertheless, the shortcut from coding to retrieval was very flexible and useful for the intermediate phase during which annotation categories had to be tested and adjusted, new markup had to be added, errors corrected etc., and a quick link into R was needed. The JAPE script had the advantage of saving quite a bit of overhead at this stage, before the annotated corpus could be officially released through an on-line query engine.

The procedure described here is certainly one with a steep learning curve, but it also offers maximal flexibility and, in the long run, helps a lot with those aspects of the manual annotation which can be automatised.

2.3 Linguistic Preliminaries: What is a subject?

The notion of subject is a controversial one, and some scholars have proposed to dispense with it entirely in certain domains. In the present section, I summarise shortly what is said about this notion in its broadest sense, in the 'classical' approach by Keenan (1975), as it was modernised and enhanced from a typological perspective by Givón (2001, 175ff). This is intended as a background to the discussion of null subjects in this thesis, rather than as a theoretical statement, since those kinds of subjects which will play a role in the remainder are fairly uncontroversial anyway.

Givón distinguishes (i) coding properties and (ii) behavior-and-control properties – terms originally due to Keenan (1975) – for *all* grammatical relations. Among the coding properties, he lists word order, verb agreement, and nominal morphology (case marking). Thus, e.g., Mandarin (cf. (1-a)) uses a fixed word order, but no case or agreement to distinguish both subject and direct object, whereas Early Latin (1-b) used word order, case, and – only for the subject – also agreement:[5]

(1) a. wo he-le   san   bei kafei le   [Mandarin]
         I   drink-PFV three cup coffee CRS
         'I drank three cups of coffee.'

     b. vir-ø   puer-o   libr-um   ded-it   [Classical Latin]
         man-NOM   boy-DAT book-ACC give-PAST-he
         'The man gave a book to the boy.'   (Givón 2001, 175f)

The behavior-and-control properties encompass a number of different syntactic "processes", as Givón calls them, of which the following are relevant for the relation 'subject' in English (Givón 2001, 177-189):

---

[5]The glossing is Givón's, viz. the unusual marking of the 3SG ending as "pronominal". PFV stands for "perfective", and CRS glosses intensifying clausal particles.

- raising: *Mary continued to see John.*

- raising to object: *They declared him to have won the war.*

- reflexivisation and possessive reflexives: *Mary hates herself. / The woman told the man to feed his own child.*

- passivisation: *Mary was seen by John.*

- zero anaphora: *Mary met Joe, Ø talked to him, and …*

- participial adverbials (i.e., control)

The coding properties and the behavior-and-control properties may end up in conflict in a given language. Thus, e.g., dative subjects in Spanish are not marked nominative, but can control into infinitival structures:

(2) me-gust-a   hacer-lo [Spanish]
     1SG.DAT-please-3SG-SUBJ do-it
     'I like to do it', 'it pleases me to do it'   (Givón 2001, 194)

In other languages, more such properties may be affected by distinctions in grammatical relation. Thus, e.g. Ute is shown by Givón (2001, 183) to use different strategies depending on whether the subject or an object is being *relativised*.

Keenan also lists several *functional* properties typical of subjects:

- independent existence

- indispensability

- absolute, presupposed or persistent reference

- definiteness

- topicality

- agentivity

According to Givón, the referential properties among these may all be reduced to one: *topicality. Agentivity* is erroneously placed here, because subjects may have other semantic roles than Agent Givón (2001, 196). In the approach which Givón advocates, these properties are put on a joint hierarchy with the above coding and behavior-and-control properties, resulting in

(3) a. Functional reference-and-topicality properties

b. Behavior-and-control properties

c. Word-order

d. Grammatical agreement

e. Nominal case-marking

This hierarchy is supposed to run from most universal (most transparent) (3-a) to least universal (least transparent) (3-e). The higher properties on the scale are also the more prototypical ones for subjects. Interestingly, we now find nominative case – the defining feature for subjects in a traditional, "Eurocentristic" view – lowest on the scale. But the scale in (3) is mainly a theoretically motivated programme so far; we would now expect massive typological evidence in its support – showing, e.g., that grammatical processes are driven by topicality rather than by agreement or case-marking.

3 Referential Null Subjects

3.1 Introduction

The omission of coreferential subjects is a classical topic of Slavonic linguistics, with earliest remarks on Ru going back at least to Lomonosov's *Rossijskaja grammatika* (1755). Lomonosov states the following rule:

> "ja, ty, on" *so svoimi množestvennymi mogut" byt' umolčany pered" vremenami glagolov", kotorych" lica okončanijami različajutsja, čto k" ukraščeniju i k" važnosti služit" [...] Kogdaž" lica okončanijami ne različajutsja; ne dolžno on yx" městoimenij umalčivat': razvě gdě iz" predšedšago razuma lice poznat' možno po bespereryvnomu i bliskomu razumu"*

> '*I, you, he* with their plurals may be elided before those verbal tenses which distinguish for person by their endings, which enhances decoration and importance [...] But when [verbs] do not distinguish person by their endings, those pronouns must not be elided, unless person can be recognised from the preceding sentence by an non-separated and close sentence' (Lomonosov 1755, 200f)

Three aspects of this rule are of importance for us. Firstly, it relates subject pronominalisation to verbal person marking. Secondly, it relates subject pronominalisation to the accessibility of an antecedent in the syntactic environment. And thirdly, it normatively prescribes, for mid 18th cent. Ru, subject realisation as the default case, and subject elision as a stylistically marked option for high registers.

The idea of a relation between subject pronominalisation and verbal person marking can actually be traced back a lot further – to ancient Greek grammar (Apollonius Dyscolus, cf. Roberts and Holmberg 2010, 3) – and will return below in a revised and improved form as the *verb impoverishment hypothesis*. The context-dependence of null subjects is of primary importance for less grammar-centered, more discourse-oriented analyses. Both aspects are useful in cross-Slavonic comparison. The third, language-specific point directly relates to diachronic change of null subject properties, our main topic. It indirectly indicates that coreferential null subjects had become optional and stylistically obsolete in 18th cent. Ru, whereas subject pronominalisation was "modern" and progressive, though possibly not yet fully established. We hope to shed light on this indirect impression by scrutinising the diachronic situation in Ru and contrasting it with developments in Pol and Cz, two closely related canonical null subject languages.

The chapter is organised as follows: After presenting the basic pattern arising from the data (section 3.2), I first review current theories of referential null subjects and their ramifications (section 3.3). Summarising the major available analyses, I concentrate on their partitioning

of the evidence rather than on technical details, and especially point out statements about language typology and language change. A short characterisation of the synchronic situation in Ru, Pol, and Cz follows in section 3.4. I then turn to the main body of the chapter, namely, diachronic change in the realm of referential null subjects in Ru, Pol and Cz (section 3.5). Following a discussion of previous analyses in the literature, a twofold evaluation of the available corpora is presented: an analysis of null vs. pronominal subjects in samples of narrative texts over time, and an analysis over genres in 16th-17th c. Ru. Qualitative aspects include changes in the licensing contexts for null subjects. The conclusions are presented in section 3.6.

3.2 Referential null subjects in Northern Slavonic: The basic picture

All Slavonic languages are null subject languages in the sense that fully grammatical clauses lacking an overt subject can be found in corpora and are readily accepted by native speakers. This has been illustrated in chapter 1 by an example from Bulgakov's *Master i Margarita*, repeated below:

(1)   a.   And this the secretary was unable to imagine, though <u>he</u> knew the procurator well.

    b.   Und das vermochte er sich nicht vorzustellen, obwohl <u>er</u> den
but that could-M.SG he SELF-DAT not imagine although he the
Prokurator gut kannte. [Ger]
procurator well knew

    c.   I ètogo sekretar' prestavit' sebe ne mog, xotja <u>Ø</u> i
and this-GEN secretary-NOM imagine SELF-DAT not can-PT.M.SG although also
xorošo znal prokuratora. [Ru]
well know-PT.M.SG procurator-ACC.SG

    d.   Ale nie mógł sobie tego wyobrazić, choć <u>Ø</u> nieźle
but not can-PT.M.SG SELF-DAT that-GEN imagine although not-bad
znał procuratora. [Pol]
know-PT.M.SG procurator-ACC.SG

    e.   Co však následovalo, tajemníka ohromilo, třebaže <u>Ø</u> Piláta
what-NOM however follow-PT.N.SG secretary-ACC bewilder-PT.N.SG although P.-ACC
dobře znal. [Cz]
well know-PT.M.SG

(ParaSoL)

By contrast, the typical non-null subject language English does require overt subjects and only shows phenomena of subject ellipsis (so-called "diary drop", see 3.3.2.2) in main clauses under narrowly restricted pragmatic conditions. Thus, null subjects in this type of language form a definable class, and do not just occur anywhere with some (low) probability. For this reason, we depart from the assumption that the null subject property has a grammatical, discrete core, rather than being just a matter of degree or even of a continuous cline.

On the other hand, the Slavonic languages do not behave equally with respect to null subjects. The typical pattern is that (standard) Ru realises a pronominal subject in contexts where Cz and Pol do not:

(2) a. Styopa finally recognized the pier-glass and realized that <u>he</u> was lying on his back in his own bed

    b. Endlich erkannte er den großen Spiegel und begriff, dass <u>er</u> rücklings
        finally recognised he the big mirror and realised that he backwards
        auf seinem Bett lag [Ger]
        on his bed lay

    c. Stepa nakonec uznal tjurmo i ponjal, čto <u>on</u>
        S.-NOM finally recognize-PT.M.SG mirror-N.ACC.SG and understand-PT.M.SG that he
        ležit navznič' u sebja na krovati [Ru]
        lie-3SG backwards at self-GEN on bed-PRP.SG

    d. Stiopa rozpoznał wreszcie tremo i zrozumiał, że Ø
        S. recognize-PT.M.SG finally mirror-N.ACC.SG and understand-PT.M.SG that
        leży na wznak we własnym łóżku [Pol]
        lie-3SG backwards in own-PRP.SG bed-PRP.SG

    e. Rozeznal vysoké třídílné zrcadlo a došlo
        recognize-PT.M.SG high-ACC.SG three-folded-ACC.SG mirror-ACC.SG and occurr-PT.N.SG
        mu, že Ø leží naznak ve své posteli [Cz]
        him-DAT that lie-3SG backwards in self's-PRP.SG bed-PRP.SG

                                               (ParaSoL)

While Cz and Pol behave like canonical null subject languages, the status of Ru is less clear: It usually resorts to subject pronominalisation as in (2-c), but it can also leave the subject out in configurations in which a non-null subject language would obligatorily realise it (cf. (1)).[1] This state of affairs has lead to controversial treatments of Ru in the literature (see below). From a diachronic perspective, Ru has been changing its status over time, or at least redefined which of its varieties contributes to the standard in the given domain. OCS and early Old Ru used null subjects in contexts where an overt subject pronoun would nowadays

---

[1] The case of embedded clause subjects is chosen on purpose in the above examples, because it is less influenced by discourse factors and brings out the grammatical distinctions more clearly (cf. also Roberts and Holmberg 2010).

be obligatory:

(3) a. Се слышавъ   Левъ   въскорѣ   посла   я₁,   и   Ø₁
         this hear-CVB.PT.ACT L.-M.NOM.SG immediately send-AOR.2SG them and

         придоста   ко цареви₂, и   Ø₂ рече   има: [...] [Old Ru]
         come-AOR.3DU to emperor and   say-AOR.3SG them-DAT.DU

     b. [Услышав   об   этом, Лев вскоре   же послал   их,] и
         hear-CVB.PT.ACT about this L. immediately MP send-LPT.M.SG them and

         пришли   они к царю, и   сказал   он им: [...] [Ru]
         come-LPT.PL they to emperor and tell-LPT.M.SG he them

     'When Lev had heard about this, he sent for them immediately, and they came to the
     emperor and he told them: [...]'   (LavrLet; trl. Lichačev)

The above quote from Lomonosov (1755) suggests that by mid 18th c., null subjects had come to be perceived as belonging to this old-fashioned, formal and literary ("decorated") register – which is interesting in itself, because nowadays they are (generally speaking) rather associated with sloppy, colloquial style. The question which we address here is when and how this change in null subject licensing took place, and what its ramifications and correlating phenomena were.

Cz and Pol, which seem to have been canonical null subject languages throughout their history, will mainly provide a baseline for comparison in the present chapter. In the synchronic domain, their licensing contexts for null subjects differ from those of Ru in systematic, describable ways. In diachrony, Cz and Pol preserved explicit verbal person marking in all tenses. Changes in the meaning and use of overt 3P subject pronouns did take place, however, and will be analysed below. Pisarkowa (1984) claims that Old Pol essentially became a null subject language only in 15th century, while subject pronominalisation was the rule earlier. We will address the empirical side of this claim below, arguing that the apparent contradiction with the development in Ru cannot be held up.

3.3 Referential null subjects in linguistic theory

3.3.1 Null subjects as a contextual phenomenon

The approaches presented in this section view referential null subjects as an instance of the more general phenomenon of *ellipsis*. Ellipsis is the omission of elements (of any kind) that are recoverable or inferable from the context (Crystal 1985, 107f), and *pro* fits this (very broad) definition, as we shall see. Analyses in this tradition try to uncover under which conditions *pro-drop* vs. subject realisation as a pronoun or lexical NP is actually chosen in

a given language. Given that in historical linguistics, we depend on corpus evidence, i.e., a specific form of language use, it is important to know these conditions and estimate their relevance. Following Huang (2000), at least four theoretical models may be distinguished: Relevant factors discovered by proponents of (i) include the linear distance (in sentence or clause boundaries) and the number of intervening further referents between antecedent and coreferent element, and the status of the antecedent as a theme or protagonist. (The latter also covers factors like animacy, agentivity, and frequency of reference.) Theories in group (ii) stress the relevance of boundaries of textual units, such as turn, paragraph, or episode breaks. Theories in (iii) and (iv) mainly add explanatory frames; an empirical phenomenon which they address is the relevance of repetitions and parallel structures for reference tracking.

On the analysis based on the notion of *topic continuity* (Givón 1983), factors like

- linear distance (in the sense of number of clause or sentence boundaries) between the antecedent and the anaphoric expression

- number of mentioned further referents between the antecedent and the anaphoric expression

- stability of the thematic status of the antecedent (maintenance or change of the protagonist)

influence the likelihood for an anaphor to be realised as a reduced expression, i.e., a pronoun or null. Referring expressions themselves may be located at different points of the so-called *topic-coding devices scale* (ordered from most accessible to least accessible – Givón 1983):

(4)   zero anaphora > unstressed/bound pronouns or grammatical agreement > stressed / independent pronouns > R-dislocated DEF-NPs > neutral-ordered DEF-NPs > L-dislocated DEF-NPs > Y-moved NPs ('contrastive topicalisation') > cleft/focus constructions > referential indefinite NPs

The higher an item on this scale, the more it is linked to a high level of topic continuity, i.e., a short distance towards the antecedent, with few or no intervenors, and continous maintenance of the protagonist status of the antecedent. Protagonists in the sense of this model are usually agentive and animate, prominent in the story by being named and referred to frequently, get introduced in the beginning and later occur in several scenes (Huang 2000, 154, citing McGann and Schwarz 1988). While Huang (2000) stresses the high amount of cross-linguistic evidence in favour of the topic continuity model, he also criticises it as insufficient

in some cases: E.g., in question/answer pairs in Chinese, lexical NPs are regularly used in order to refer to an antecedent introduced in the immediately preceding turn, i.e., in a situation of high topic continuity. On the other hand, English pronouns (= highly accessible items) can occur after long stretches of text with clearly interfering referents, in so-called "return pops" (Huang 2000, 156). The latter phenomenon can be better addressed by the *hierarchy model* (ii) (Fox 1987): According to the latter, lexical NPs are used at the beginning of a structural unit of discourse, such as a turn, paragraph, or episode, whereas reduced anaphoric expressions refer to lexical NPs within the same unit. Conversely, the use of a coreferent pronoun can also signal the intended continuation of a structural unit which was about to be closed. This seems to be the case in "return pops"; if a lexical NP were used here, the structural unit would have been understood as closed and a new one started. (iii) A third approach tries to tie the findings of the first two to *cognitive factors* such as memory and attention. Its central claim is that a full NP is used when the respective referent is expected not to be currently activated in the memory of the addressee, whereas a reduced anaphoric expression is applicable if its referent is thought of as currently activated both for the speaker and the addressee (Gundel and Zacharski 1993; Kibrik 1996). The amount of activation of a referent may be raised by explicit mention, by paying attention to the extra-linguistic environment shared by speaker and hearer, or by retrieving encyclopaedic knowledge on the referent from long-term memory. Depending on the language, different forms may be used to express different levels of activation on a universal *Givenness Hierarchy* (Gundel and Zacharski 1993) which reaches from *in focus*[2] down to merely *type identifiable*. The higher the degree of givenness of an entity, the more likely it is to be realised by a reduced form (pronoun or zero).

Kibrik (1996) stresses that a cognitive account is clearly more adequate than (i) and (ii), since it is utterly implausible that speakers "look back in the pre-texts and trace antecedents there" (p. 256). On his view, the decisive coginitive factor for the use of a coreferential pronoun is presence in the speaker's active memory (activation), which a referent can enter from previous discourse, observed environment, or long-term memory. Once a referent is in active memory, stable semantic properties such as animacy or "protagonisthood" can also influence its current activation level, with protagonisthood being relevant mainly for reactivation (p. 271).

However, as Huang points out, the correlation between discrete grammatical means and the continuous cognitive hierarchy is only imperfect: Sometimes, pronouns have to be used to

---

[2]To avoid a potential misunderstanding, *focus* designates a cognitive status here, by no means equal to focus in the sense of the linguistic theory of information structure. It is exactly the entities in (cognitive) focus which are most likely elided, and thus, belong to the background, in the linguistic sense. Conversely, pronouns in focus in a linguistic (information-structural) sense, are never elided.

clarify an anaphoric link to one of a set of equally activated referents (Kibrik 1996, 300 for Ru); in other cases in Chinese, pronouns can refer to an unactivated entity (Huang 2000, 163). Huang himself opts for (iv) a neo-Gricean *pragmatic model.* On his view, the distribution of elaborated vs. reduced forms follows from an interplay of conversational maximes such as not to say less than necessary to be understood (Q-maxime, Levinson 1983), and not to say more than is required (I-maxime, Levinson 1983). The most compelling argument for a treatment on the pragmatic level comes from conversations showing anaphoric repair strategies, in which the speaker makes the intended coreference understood to the hearer by stepwise relaxation of the I-maxime. Interestingly, the linguistic devices marking anaphoric repair generalise to other types of conversational repair as well. Furthermore, Huang presents interactional explanations for the problematic cases mentioned above. E.g., "return pops" are typically marked by repetitions of key lexical items from the sentence containing the antecent, or by structural parallelism.

Not all of the empirical factors found to be relevant by these four approaches to discourse anaphora may be operationalised for a corpus-based diachronic study. Gathering the most important points, at least the following might be relevant:

- linear distance between antecedent and anaphoric expression (number of clause/sentence boundaries)

- intervening other referents

- type of expression used for both anaphor and antecedent (zero form, pronoun, lexical NP etc.)

- semantic properties of the antecedents which influence its saliency: animacy, agentivity; number of previous and overall mentions

- text-structural units: paragraph and episode breaks

- structural and lexical repetition/parallellism between the environments of the antecedent and the anaphoric expression

Some of these – number of intervenors, number of previous and overall mentions, paragraph breaks – can be read off the corpus automatically once all coreference links have been annotated. Others – animacy, type of expression – can be derived from morphosyntactic tags. The factors which do require special manual annotation are agentivity of the antecedent and structural repetition/parallelism. Unfortunately, most empirical claims of the cognitive

model (iii) and the pragmatic approach (iv) have to go unconsidered. (iii) requires other (psycholinguistic) types of data; and the most compelling arguments for (iv) depend on conversational interactions which (narrative) historical texts often lack. Finally, it should be noted that cross-linguistic differences come out as mere matters of degree on the above-mentioned accounts. I.e., they have little to say about the question whether a specific language uses a pronoun or a zero form for discourse anaphors, apart from the claim that both are reduced forms compared to a lexical NP, and depend on the above factors.

Syntactic accounts usually minimise reference to discourse conditions for pro-drop. But there are good arguments *against* this strategy, at least for diachronic data: First, as McShane (2009, 103f) rightly points out, some syntactic treatments (she explicitly refers to Lindseth 1998) involve the factor of stylistic markedness anyway, which can only be judged on the basis of a contextual model. Secondly, it is obvious that pro-drop languages do not principally disallow overt pronominal subjects, but merely restrict their use to certain discourse and/or information-structural conditions, which are more narrowly defined than in non-pro-drop languages. Working with historical texts, we depend on the empirical manifestations of those conditions. Thirdly, research by Nichols (1984), Kibrik (1996) and Lindseth (1998) on Russian, McShane (2009) on Russian and Polish, and Kresin (1994) on Russian and Czech has shown that there is a "grey area" in these languages, in which both pronouns and null elements may be used. Without taking into account the system of anaphora licensing in discourse as a whole, we would have no means to treat such examples. Note that these points apply irrespectively of whether we wish to view null subjects as a gradual, continuous matter or as parametric: We need to take discourse conditions on coreference relations into account. Nevertheless, as we will argue in 3.4, referential null subjects in Slavonic languages cannot be captured adequately by *only* analysing their discourse conditions.

3.3.2 Null subjects as a grammatical phenomenon

3.3.2.1 Pro drop: Parametric vs. typological approaches

Theories of a different kind treat null subjects as a phenomenon belonging to the language system rather than to the realm of usage. Although some of these approaches do not dispute the fact that there are important pragmatic, semantic, and text linguistic conditions which influence the omission of subjects in actual utterances, the tacit assumption is that the null subject property has an irreducible grammatical core. One of the empirical arguments for this view is that the omission of subjects in a non-null subject language [(non-)NSL] generally leads to strong grammaticality violations which cannot be remedied by non-grammatical

means (e.g., discourse factors), while the class of exceptions to this rule is restricted and definable. We have seen an illustration of this fact for embedded clauses in (2) above. An important conceptual argument for the grammar-centred view is that the null subject property correlates with other grammatical phenomena, so that a unified explanation of both becomes possible by the assumption of one common rule. This type of argument has played a chief role in the framework of Chomskyan generative grammar, which views the null subject property as a paradigm example of a *parameter* of Universal Grammar (UG) (Haider 2001). For this reason, the theoretical literature abounds – which will not be fully reflected here, because discussions of the history of research on the topic are readily available (cf. especially Müller and Rohrbacher 1989 and Kosta 1990 for the earlier government-and-binding tradition, and Roberts and Holmberg 2010 for recent minimalist approaches).

Parameters are abstract features of a grammatical system in the form of binary oppositions; setting a parameter to "+" or "−" amounts to allowing or excluding a whole cluster of syntactic properties. Within the Chomskyan framework, they are understood as a sort of switches controlling the language-specific expression of the human language faculty: Exposition to certain pieces of key evidence leads the language-acquiring child to setting a parameter in one way or the other, which in turn determines properties of his/her intuitive knowledge about the mothertongue. But independently of UG, parameters may also be viewed as explanatory devices intended to capture typological variation between languages. They make clear and testable cross-linguistic predictions about clusters of syntactic properties. Even if those predictions often turn out as too coarse, leading to the postulation of subparameters, microparameters etc., this is a great heuristic advantage. Recently, the nature and explanatory potential of UG parameters has been criticised sharply by Newmeyer (2004), among others with reference to pro-drop (cf. his debate with Roberts and Holmberg – Newmeyer 2004, 2006; Roberts and Holmberg 2006; 2010).

The term *pro* drop dates back to transformational syntactic analyses which postulated a deletion operation on pronominal subjects (Perlmutter 1971). Its connection to the pronominal system seems obvious – null subjects in *pro* drop languages would typically be realised as pronouns in a language like English. Already the case of closely related German shows, however, that not all pronominal subjects are equally droppable: German allows for the absence of certain expletive subjects, but not for the absence of subjects which realise a thematic role (cf. Grewendorf 1990 for details). The reverse – a language which requires expletive subjects, but drops argumental subjects – has not yet been attested. A number of syntactic properties have been related to the null subject or "pro[noun] drop" (Perlmutter 1971) parameter, in search of a unified explanation. The "pro-drop syndrome" (Haider 2001, 285) as postulated

by Rizzi (1982; 1986) lumps together

(5) a. referential null subjects

     b. the absence of subject expletives

     c. the absence of *that*-t effects

     d. the possibility of postverbal subjects (in VO languages)

(5-a-c) can already be traced back to Perlmutter (1971), who had furthermore assumed that only non-null subject languages can have true arbitrary subject pronouns (rather than, e.g., clitics), such as English *one* (Roberts and Holmberg 2010, 3). For completeness, we may add a further factor which figures separately in many analyses, namely, the properties of verbal agreement marking. We will deal with this factor separately in 3.3.2.3.

A typical non-pro-drop language like English shows neither of these phenomena: It disallows referential null subjects (6-a), requires subject expletives (6-b), shows *that*-trace effects – i.e., a ban on extraction from an embedded subject position if a complementiser is present – (6-c), and prohibits postverbal subjects (except in special configurations; (6-d)).

(6) a. *Speaks English.

     b. *Rains.

     c. *Who did you say that ˍ wrote this book?

     d. *Have telephoned many students.

A typical pro-drop language like It[alian] behaves the opposite way: It allows referential null subjects (7-a), needs no subject expletives (7-b), does not know *that*-trace effects (7-c), and freely uses postverbal subjects (7-d):

(7) a. Parla inglese. [It] (=(6-a))

     b. Piove.[It] (=(6-b))

     c. Chi hai detto che ˍ ha scritto questo libro? [It] (=(6-c))

     d. Hanno telefonato molti studenti. [It] (=(6-d)) (Roberts and Holmberg 2010, 16ff)

Haider (2001) reckons that the predictive value of the pro-drop parameter in its present shape has been rather disappointing: As Gilligan (1987) has shown, only the existence of referential null subjects and the absence of subject expletives are connected without exception cross-linguistically, but (5-a) and (5-c) or (5-a) and (5-d) are not. E.g., Finnish and Georgian have (5-a), but not (5-c); several languages have (5-a), but not (5-d) – cf. also Müller and Rohrbacher (1989). According to Haider, this is not too unexpected, since the theoretical

explication of the link between the properties in (5) was unconvincing (p. 286). On his view, null subjects are influenced by various grammatical circumstances: null subjects may conform to constraints on clitics (Italian) or topics (Chinese); some languages have null subjects in general (Italian), some only for expletives (Icelandic). The consequence of Haider's view then seems to be that pro-drop should not be taken as a unitary phenomenon driven by a single parameter setting, but rather an epiphenomenon of the functioning of various independent parameters.

Newmeyer (2004) takes the critique of the pro-drop parameter much further: For him it is but an instance of the general inadequacy of a parameter-setting approach to typological variation. He once more refers to Gilligan (1987), remarking that of the four implications he found valid in a 100 languages sample (cf. (8)), three are just due to the fact that languages with overt expletives are so rare, i.e., to a sampling error; this leaves (8-c) as the only valid implication, a "not very heartening" result (Newmeyer 2004, 204).

(8) a. Referential null subjects $\Rightarrow$ null expletive subjects
     b. Free subject inversion $\Rightarrow$ null expletive subjects
     c. Free subject inversion $\Rightarrow$ no *that*-trace effects
     d. no *that*-trace effects $\Rightarrow$ null expletive subjects

This evaluation is not shared by Roberts and Holmberg (2006), who point out that Rizzi's original claims had been based on theoretical considerations and a very small sample of closely related languages. Gilligan (1987) confronted these claims with a massive amount of new phenomena, without always controlling for the additional factors which might further influence the effects of the null subject parameter. In this situation, it could already be seen as a success that Rizzi's central generalisation in (8-c) held up at all (this point is elaborated upon in Roberts and Holmberg 2010, 21f). The alternative proposed by Newmeyer, i.e., language-specific rules, would instead fail to make cross-linguistic predictions altogether. In a rejoinder to this critique, Newmeyer (2006) points out that an intersection of Gilligan's sample with the data on verbal person marking in Haspelmath and Comrie (2005) leads to even more discouraging results: 12 of 22 languages with null expletive subjects in the sample also show person marking on the verb, but the other 10 do not. This seems devastating for the traditional idea of a connection between explicit person marking on the verb and the availability of null subjects. We will see below in 3.3.2.3 how this issue has been addressed before by proponents of the parametric approach. Roberts and Holmberg (2010, 24ff) latest reply in this debate takes issue with the idea implicitly advocated in Newmeyer (2006) that typological distinctions should be easily ascertainable by superficial inspection of relevant

data, rather than being theory-dependent.

Null subject properties of Slavonic languages provide a good illustration of the dilemma: There is considerable disagreement in the literature about whether Ru should count as a null subject language or not (see 3.4). According to the typology of Dryer (2011), Ru has "obligatory pronouns in subject position", just like English, French or German. By contrast, Cz realises "subject affixes on the verb", and Pol shows "subject clitics on variable hosts". The criterion for inclusion in the first category is that a given language always or at least predominantly uses overt subjects. "For the purposes of this map, if all sentences with pronominal subjects on a couple of pages of text in a language have a pronoun in subject position, the language is coded as being of the first type." (Dryer 2011). As will be discussed below, the outcome of such a procedure for Ru will strongly depend on the register from which the text is taken; but even if we restrict ourselves to the standard language, there are describable syntactic configurations in which pronominal subjects will be systematically optional (see (1-c) vs. (2-c) above) and others, in which they are not. For advocates of the position that Ru is *not* a null subject language, a lot hinges on isolating (and possibly, "explaining away") those retreats of optionality. In any event, it is rather unlikely that the "couple of pages" alluded to in the above quote from Dryer (2011) will not contain null subjects.

Summarising this section, it is undeniable that the parametric approach to pro-drop could not live up to its – or rather, Rizzi's (1982; 1986) – high expectations when confronted with data from *really* diverse languages. Roberts and Holmberg (2006; 2010) may be right if they suspect that in some cases, alleged counterexamples are just a by-product of too superficial analyses. Given the difficulties which many serious approaches had at classifying Ru as a null subject language or not (cf. section 3.4.2), the utterly naive detection procedure given by Dryer (2011) is very unlikely to produce any reliable results. Still, the number of counterexamples in the literature is high, and they probably will not all be due to wrong analyses. That some – far from obvious – generalisations did hold up is a certain limited success. On our view, the right way to proceed would be simply to revise and refine the proposed parameters so that they better account for the relevant data. This can in turn provide new and better starting hypotheses for typological studies. Needless to say, there is no way around valid typological generalisations, which by any means must become part of the modelling. We will see that the parametric theory of null subjects makes useful and testable predictions also for the diachrony of Slavonic (and other) languages.

3.3.2.2 **Pro-drop and Non-pro-drop**

Within the parametric approach, a number of problematic cases of null subjects in otherwise clear non-null subject languages have been identified. The (unpredictable) occurrence of such missing arguments would of course undermine the [±pro-drop] divide, and ultimately, the parametric view on null subjects.

German shows the phenomenon of so-called *topic drop* (Ross 1982):

(9) a. $_{-i}$ Hat sich noch nicht gezeigt. [Ger]
                  has SELF yet   not   shown
            '(He) hasn't shown up yet.'

     b. $_{-i}$ Habe ich schon t_i gesehen. [Ger]
                  have   I   already   seen
            'Have seen (it) already.'   (Haegeman 1997, 240)

Topic drop is restricted to the leftmost position of root clauses, and elides an element exclusively from the so-called *Vorfeld (pre-field)*, leaving a verb-first clause. Such restrictions are unknown in cases of pro-drop. Furthermore, topic drop applies to any kind of argument, not only to subjects (cf. topic drop of the object in (9-b)).

Haegeman (1997) analyses dropped subjects in English and French. Her findings, in short, are the following: Dropped subjects only occur in very restricted registers: diaries, informal notes, commercial packaging, spoken utterances in certain colloquial idiolects, cf. (10-a). There is no evidence for the concomitant features of the "pro-drop syndrome" (e.g., lack of *that*-t effects, free subject inversion, rich verbal agreement) in their environment, except possibly for null expletive subjects (10-b).

(10) a. Contains carotene. (Carrot Facial Oil, The Body Shop)
      b. Rained in the night, wind, rain and hail.   (Haegeman 1997, 237/240)

"Diary drop" – as it came to be called – only targets left-peripheral elements, and exclusively subjects of root clauses, according to Haegeman (1997). As with topic drop, these are restrictions unheard of with pro-drop of the Italian type, so the evidence in (10) should not lead to a revision of the parametric analysis of pro-drop as a whole. Haegeman presents an analysis building on Rizzi (1982; 1986; 1997), according to which *pro* can exceptionally be licensed in exactly the position to which diary drop in fact applies. This is independent of the fact that *pro* cannot be "identified" (see 3.3.2.3 on this terminology) by verbal agreement in English.

In a paper based on a (still) larger corpus of diaries and similar texts, Haegeman and Ihsane

(1999), note that the characterisation of diary drop in Haegeman (1997) had actually been too restrictive: Although the ban on dropped arguments in any other position but the left periphery of root clauses conforms to native speakers' intuitive judgments (and has been reiterated in the literature, cf. Holmberg 2010, 90f), there are in fact rare attested cases of subject drop in embedded clauses, cf.

(11)   Understand where _ have gone wrong.   (Haegeman and Ihsane 1999, 135)

It is obvious that this empty category does not lend itself to the explanation in Haegeman (1997), which was tailored to the leftmost position of root clauses. Without arriving at a final analysis, Haegeman and Ihsane (1999) tentatively propose that speakers for whom embedded null subjects as in (11) are possible, in fact allow for a rule of pronoun deletion in their grammar, i.e., relax the conditions on identification of *pro*.

The conclusion to be drawn here is that even if English has some very restricted null subject registers for some speakers, the phenomena of "diary drop" and "topic drop" show restrictions which are not characteristic of pro-drop in clear null subject languages. Therefore, they do not undermine the parametric theory of pro-drop as a whole, but rather fall out of its realm and require a different treatment.

3.3.2.3 Verbal Agreement

The idea that in some languages, the subject of a finite clause may remain unrealized because its relevant grammatical features are displayed on the verb, is a traditional one, mentioned e.g. already by Jakobson (1971 (1939)), by Lomonosov (1755) – as we saw above – and in fact much earlier: Roberts (2007, 22 n.16) cites the grammarian Apollonius Dyscolus (2. century B.C.) with the statement

> "The nominative [subject] is implicitly present in [finite] verbs, and it is definite in the first and second persons, but indefinite in the third because of the unlimited number of possible referents." (Apollonius Dyscolus: On Syntax, Book 1, §17; Householder 1981, 25).

(Cf. also Kosta 1990 and references therein for the history of research.) The tight connection between subject realisation and verbal agreement has been adopted, in one form or other, in the most diverse analytical frameworks, from Government and Binding Theory (Taraldsen 1978) to Role and Reference Grammar (Van Valin and LaPolla 1997, 34 with respect to Croatian). In a somewhat simplified form, the hypothesis may be stated as follows:

(12)   *Inflectional Richness*:

The availability of referential null subjects depends on the richness of verbal inflectional marking.

It has been known for long that, in its simplest form, *Inflectional Richness* is blatantly wrong: There are null subject languages such as Chinese or Japanese which never signal the value of agreement categories on the verb.[3] This has lead to several refinements and elaborations of the original idea.

The set of relevant agreement features – the so-called ϕ-features – comprises person, number, and gender, as well as honorific properties (in some languages). Chomsky (1982) locates an abstract representation of these features in the structural node INFL and assumes that the phonetically empty pronominal *pro* needs to be governed by a sufficiently strong INFL, in a technical sense.

According to Rizzi (1986), *pro* must be (i) *licensed* and (ii) *identified*. It is licensed if governed by a case-marking head, such as INFL or – under additional conditions – the Comp node. It is identified if the values of its ϕ-features are determined by INFL, an option which is subject to cross-linguistic parametrization. The latter idea, namely that ϕ-feature values are transferred onto the null subject from a functional head in the verbal domain, has survived in several more recent analyses, but is challenged e.g. by Holmberg (2005, 537f).

According to Jaeggli and Safir (1989, 29f), *pro* drop can only obtain in languages with a *uniform* verbal paradigm. A uniform paradigm is characerised by distinct marking of either all of its members (as, e.g., in Spanish) or none of its members (as, e.g., in Chinese). Albeit empirically more adequate than its predecessor, this disjunction of conditions does not seem very attractive on a conceptual level.

Müller (2006; 2007) develops a theory of *pro* drop based on the general idea in (12) and an articulate theory of morphological derivation. He proposes not to take syncretism or Inflectional Richness on verbal categories at face value, but to consider its morphological derivation. The respective generalisation on the relation between *pro* and verbal agreement – located in the functional head T (the theoretical successor of the Infl node) – then reads as follows:

(13)   *Pro Generalization:*   (Müller 2007, 2)

An argumental pro DP cannot undergo Agree with T if T has been subjected (perhaps vacuously) to person feature neutralizing impoverishment in the numeration.

---

[3]Chomsky (1981, 284, fn 47) already states that the latter do not exhibit pro-drop proper, because they freely omit pronouns of different syntactic functions.

Examples of impoverishment include German verb forms, for which, e.g., 1.PL and 3.PL are always syncretic, or Icelandic verb forms, which show a systematic syncretism of 2.SG and 3.SG. Müller shows in detail how these syncretisms may be derived via rules at the syntax/morphology interface which remove the person features [±1] (German) and [±2] (Icelandic), respectively. Consequently, neither German nor Icelandic allow for *pro* drop . His account revises the general idea of Inflectional Richness, making it more precise and theory-dependent. At the same time, it overcomes the stipulativity inherent in Jaeggli and Safir's (1989) theory. Müller (2007, 11ff) scrutinises syncretism patterns in several typologically diverse languages, which had been adduced as problematic for Inflectional Richness by Cysouw (2001). It turns out that the fact that pro-drop may freely occur despite obvious syncretisms in the person paradigms, a crucial argument against Inflectional Richness, can be explained if the issue of impoverishment during morphological derivations is taken into account.

Ru presents an interesting and potentially problematic case for Müller's (2006) account: The neutralization of person distinctions in Ru past tense forms looks like a paradigm case of impoverishment; therefore, the prediction would be that Ru must be a non-null subject language. However, Müller (1988) himself (like Růžička 1986a, Kosta 1990 and others) had already argued that Ru was indeed a *pro* drop language (see 3.4 for the language-specific details), and Müller (2006; 2007) subscribes to this view. The explanation offered by Müller (2006; 2007) is that historically, the neutralization of the category of person in Ru past tense forms did not involve a morphological operation on the main verb, but the loss of a separate morphological entity, the auxiliary. Synchronically, the assumption is, then, that past tense T^0 in Ru never contains any person features, neither in the lexicon, nor in the syntax. Thus, no operation of impoverishment has to apply, and "the analysis given for pro-drop in Russian past tense contexts is identical to that given for pro-drop in languages like Korean (Japanese, Chinese)." (Müller 2007, 8)

3.3.2.4 *Pro* as a Null Pronoun

The analyses introduced in 3.3.2.3 assign a major role to the agreement features present in T^0, but they nevertheless presuppose that an empty category occupies the subject slot. Several more radical approaches have rejected this idea and proposed that agreement features on the verb can be interpreted directly, without any need for a null subject (cf. e.g. Borer 1986; Alexiadou and Anagnostopoulou 1998). Holmberg (2005) analyses partial pro-drop and subject expletives in Finnish, arguing that the evidence contradicts the latter type of analysis and forges the view that there is indeed an empty pronoun *pro* in subject position.

A surprising empirical fact demonstrated in Holmberg (2005) (first analyzed in Holmberg

and Nikanne 2002) is that Finnish challenges the generalisation that NSLs cannot have subject expletives (see above 3.3.2.1). The reverse case – a non-NSL which allows for the absence of certain expletive subjects – is attested by German (cf. Grewendorf 1990 for details); but, as Haider (2001) puts it, the fact that NSLs do not possess subject expletives, might even be regarded as the *only* generalisation about correlates of *pro* drop that has proven to hold up cross-linguistically.

Fin[nish], according to Holmberg, is an optional (or *partial*) NSL which systematically allows for the absence of 1P and 2P subjects, as well as for the absence of 3P subjects which are coreferent with a close antecedent. Furthermore, verb-initial sentences are generally ungrammatical; either some element has to be preposed to the left of the verb, or the expletive subject *sitä 'it-PART'* must occupy the preverbal slot. On the other hand, expletive subjects in sentences with meteorological predicates are ungrammatical (Holmberg 2005, 540). Finnish thus demonstrates that *certain types* of expletives may occur in (partial) NSLs, albeit with some unusual properties. Finnish *sitä*, according to Holmberg, provides evidence against the "Agr only" type of analysis alluded to above, according to which there is no null subject in the structure and the agreement features expressed in the verbal domain are interpreted directly (his "Hypothesis A", cf. also 3.3.2.3 above). The argument runs as follows: *Sitä* is disallowed in 1P or 2P clauses without an overt subject (14-a), and with an overt subject in subject position (14-b).

(14)   a. *Sitä puhun   englantia. [Fin]
            EXP   speak-1SG English

        b. *Sitä minä puhun   englantia.[Fin]
            EXP   I   speak-1SG English

        c. Sitä olen   minäkin käynyt Pariisissa.[Fin]
            EXP be-1SG   I-too   visited Paris-INE
            'I have been to Paris, too (actually).'   (Holmberg 2005, 554)

One might think that the reason for (14-a,b) were to be found in an agreement mismatch between 3P *sitä* and the verb, but (14-c) shows that *sitä* is compatible with 1P. Alternatively, *sitä* might be excluded in (14-a) because there is no need for an expletive at all in order to check for agreement. But there is no need for it in (14-c) either, and yet it may occur. The explanation offered by Holmberg ("Hypothesis B") is that a null element *pro* sits in subject position in (14-a), agrees with the verb and leaves no place in the structure for the expletive. In a nutshell, the incompatilbility of the expletive with certain subjectless sentences is only expected if it competes for subject position with an obligatory constituent of the subjectless clause (*pro*). Note that Holmberg's argument against "Hypothesis A" is somewhat indirect:

The hypothesis is not sufficient for explaining the facts. This does not necessarily mean that it is wrong – the pattern in (14) could in principle be due to further factors (but none have been proposed).

Holmberg (2005) continues to flesh out the theory of *pro*, including null subjects of generic and impersonal sentences and of embedded clauses. (We will return to this aspect of his analysis in chapter 5.) He argues that the agreement head T contains a semantically interpreted [+D](efiniteness) feature which it transfers to *pro* in a true ("canonical") NSL. In a *partial* NSL, it contains no such feature. Accordingly, the null subject of a canonical NSL can refer definitely to an individual or a group, but is incompatible with a generic reading; the null subject of a partial NSL, however, must be referentially bound by a nearby expression (usually in the same clause) or it is automatically interpreted as indefinite or generic. The "syndrome" of partial pro-drop and, in the case of a lack of referential binding, null subjects with generic or indefinite interpretation is demonstrated on Brasilian Portuguese, Marathi, and Hebrew data. As will be shown in detail in chapter 5, modern Ru also fits this picture quite well, and the correlation can even be demonstrated in diachrony.

On the empirical side, a partial NSL will use null subjects only in a morphosyntactically defined subset of cases. Thus, e.g., Hebrew has null subjects in 1P and 2P, but overt pronominal subjects in 3P (Vainikka and Levy 1999). Holmberg (2010, 91ff) substantiates the distinction between partial and consistent NSLs as follows: In the equivalent of a construction like

(15)   John said that he/∅ wanted to buy a car.

in a consistent NSL (e.g., Arabic, Greek, Spanish, Turkish, Italian), the subject of the embedded clause *must* be null, whereas it is only *optionally* null in a partial NSL (e.g., Brasilian Portuguese, Finnish, Marathi). If, in a partial NSL, the embedded subject were to refer to an entity *different* from the matrix subject, an overt pronoun would be obligatory. This is not the case in a consistent NSL; unless under conditions of contrast or shift of the current discourse topic, an embedded null subject can also corefer with an external, third entity, cf.

(16)   Gianni$_1$ non ha detto niente, ma Paolo$_2$ ha detto che ∅$_1$ vuole comprare una macchina nuova. [It]
         'Gianni$_1$ hasn't said anything, but Paolo$_2$ says that he$_{1/2}$ wants to buy a new car.'
                                                                     (Holmberg 2010, 92, (9b))

According to Holmberg (2010, 94), the typical case in a partial NSL is for definite subject pronouns to be only *optionally* null; but if they are null, they depend on a locally c-commanding antecedent.

Roberts (2007b) draws from Holmberg's (2005) and Müller's (2006) findings when developing a mixed approach, in order to deal with diachronic findings from French and Portuguese. According to him, impoverishment in T^0 (cf. Müller 2006) *precludes* the presence of a [+D] feature (cf. Holmberg 2005). This leaves two options: If there is an empty element *pro* available in the lexicon of the given language, its properties will be determined along the lines of Holmberg (2005), i.e., it will have to be coreferential with a nearby antecedent (in the same clause) or it will be interpreted as generic or indefinite. If there is no such empty element *pro* in the lexicon, null subjects will be excluded completely.

3.3.2.5 Pro-drop and diachronic change

Comparing the history of French on the one hand, and a conservative variant of Portuguese (European) and a progressive one (Brasilian) on the other, Roberts (2007) shows that Portuguese changed from a canonical NSL to a partial one, whereas French went through a partial *pro* drop phase and then lost pro-drop completely. The empirical findings reported by Roberts (2007) are summarized in table 3.1 below.

The chief reason for French moving from a partial null subject language to the current non-null subject type, [4] according to Roberts, is to be found in two concomitant developments in Middle French: the loss of verb-second (V2) movement in declaratives, and the rise of a full series of atonic, weak subject pronouns. In the present context, note that the fact that a weak subject pronoun systematically replaces *pro* makes most sense if *pro* itself has the status of a weak pronominal category.

Lehmann (1995, 13ff) analyses the correlation between null subjects, pronouns, and verbal morphology as a typical *grammaticalisation* process. He discusses the relevant changes in late Latin: Instead of the multifunctional personal pronoun *is/ea/id 'he/she/it'*, which could be used deictically or anaphorically in Classical Latin, the former demonstrative *ille* became used as a marker of anaphoricity. Lehmann analyses this step as beginning *syntactisation*: The function of *ille* is reduced to that of syntactically representing, in its clause, a previously mentioned noun phrase. The next step, *morphologisation*, is an erosion of form, typically indicated by pronominal subject doubling, such as, e.g., in languages of the Balkan. The subsequent steps, *cliticisation* and *affix formation*, lead to a typical *pro* drop language, in which all the information formerly encoded in the pronoun is now implicit in the verbal affix. The analysis assumes no empty category or null pronominal element as a remainder of these processes. According to Lehmann (2002, 35), demonstratives are a common source

---

[4]Note that French behaves like a NSL in inversion contexts, such as root interrogatives, counterfactual conditionals etc. – cf. the discussion in Roberts (2007b, 30ff).

Brasilian Portuguese	European Portuguese	French historical	French synchronic
impoverished verbal paradigm	explicit verbal paradigm	since XII. century impoverished verbal paradigm	impoverished verbal paradigm
3P null subject needs local antecedent	3P null subject takes any antecedent	ca. 1450-1600: 3P null subjects only with local antecedent	no 3P null subjects
in neutral contexts, 1/2P, but no 3P null subjects	in neutral contexts, 1/2P and 3P null subjects	ca. 1450-1600: in neutral contexts, 1/2P, but no 3P null subjects	no 1/2P null subjects
null subjects with a generic reading	no null subjects with a generic reading; instead, constructions with *se* subject clitic	(no indication)	no null subjects with a generic reading; instead, overt subject *on* or mediopassive construction
higher frequency of 3P pronouns	lower frequency of 3P pronouns	(no indication)	(no indication)
expletive subjects always null	(no indication)	ca. 1450-1600: expletive null subjects possible	expletive subjects always overtly realized

Table 3.1: Diachrony of null subjects in Portuguese and French (Roberts 2007b)

for the development of 3p personal pronouns (cf. German *er* < **ei-s*), whereas 1P and 2P personal pronouns are often derived from nouns denoting social relations. Therefore, it may not come as a surprise if the respective grammaticalization processes differ. On the other hand, Lehmann (2002, 36) notes that pronouns of all three persons in languages like German, English or Russian form a common paradigm and are morphologically maximally similar, which would speak against different pathways of change.

If we take into account the written history of French and Portuguese (Roberts 2007b and references therein), German (Abraham 1991) and Russian (section 3.5), we find that null subjects are diachronically being lost rather than established (as opposed to the Latin case discussed by Lehmann 2002). Grammaticalization should be able to account for this fact. The only reason compatible with Lehmann's analysis could be that the introduction of a new anaphoric pronoun launches a new cycle of syntactisation, morphologisation, cliticisation and affixation. However, the trigger which caused the establishment of the weak pronoun

in the first place, remains unclear.

Fuß (to appear) argues, in a similar vein as Lehmann, that null subjects historically arise via a reanalysis of weak pronouns and pronominal clitics as verbal agreement markers, and thus, systematically correlates with gaps in the paradigm of weak pronouns. If new weak pronouns are established, *pro* drop is predicted to decrease, because these more specific morphemes block the insertion of the null element (Fuß to appear, 7, discussing Finnish colloquial data from Vainikka and Levy 1999). So *pro*, in this analysis, is a weak, *underspecified* pronominal. The relation between *pro* and verbal morphology is described in essentially the same way as in grammaticalization theory (Lehmann 1995; cf. also 3.3.2.3 above).

3.3.2.6 "Radical *pro* drop "

Neeleman and Szendrői (2007) develop a theory of so-called "radical" pro-drop of the kind obtaining in languages like Japanese, Korean and Chinese, and test it on a large sample of languages, including Ru. Radical pro-drop affects subject and object pronouns. The general idea is that radical pro-drop presupposes a transparent, agglutinative morphology of personal pronouns. Fusional, intransparent morphology blocks radical pro-drop. Neeleman and Szendrői (2007, 705) discuss Ru as a potential counterexample to their theory, assuming that it is *not* a (radical) pro-drop language. However, the paradigm of 1/2P personal pronouns shows some signs of agglutinative morphology, namely, oblique endings which are homonymous to endings of the 2nd nominal declension (e.g., *mn-oj 'me-INS '*, *mn-e 'me-DAT'*). Neeleman and Szendrői (2007) argue that this amount of transparency is not consistent enough to be exploited by the language learner; e.g., it does not hold for the plural pronouns. Considering the massive options for (subject and object) ellipsis in Ru, it might still be promising to treat it as a radical pro-drop language.

3.3.3 Multifactorial approaches

The analyses of null subjects presented in this section combine grammatical and contextual aspects of pro-drop in a single model.

Largely theory-independent lists of relevant factors which rule the choice of expression between a type of prononun, *pro,* or lexical NP, have been proposed e.g. by Bresnan (1998) and Vincent (2000, 31):

(17) a. *semantic*:
          anaphoric or text-external (deictic) reference; (in)definiteness
      b. *information structural*:
          focus vs. background, topic; maintained discourse topic or topic change
      c. morphosyntactic: overt realisation of agreement categories (person, number, gender)
      d. syntactic: hosting of appositions, correlative constructions

When working with historical texts, (17-c) will certainly be most problematic, and a lot will hinge on the possibility to relate it to overt markers e.g. of topichood or contrast.

Frascarelli (2007) and Frascarelli and Hinterhölzl (2007) depart from the well-known fact that even in a canonical NSL like Italian, *pro* cannot just occur anywhere, but depends on a contextually present and accessible antecedent. Formulating the conditions on this relation, they rely on the following classification of topics:

- *Aboutness-shift* Topics encode referents which are "newly introduced, newly changed or newly returned to" (Givón 1983). This is intended as a stronger requirement than just pragmatic "aboutness"; there can be only one Aboutness-shift topic per sentence.

- *Contrastive* Topics induce semantic alternatives, without changing the focus value of their sentence (Büring 1999). Again, there can be only one contrastive topic per sentence.

- *Familiar* Topics represent given, accessible referents (= background material) and signal topic continuity in discourse. There can be several Familiar Topics per sentence, which are typically destressed and realised in pronominal form (Pesetsky 1987).

Continuing Topics (Givón 1983) are no separate type, but simply Familiar Topics which corefer with the closest Aboutness-shift Topic. The contextual condition on the identification of *pro* then says that *pro* must be coreferent with the *closest Aboutness-shift Topic*. An intervening, different Aboutness-shift Topic would disrupt this relation and demand coreference with *pro* itself. Since the contextual and information-structural configurations at stake in Italian are quite similar to those in Slavonic languages, we will illustrate them with Slavonic examples in 3.4 and only point out some facts discussed in Frascarelli (2007) in order to clarify the above condition. First, constituents which are not topics do not qualify as antecedents of *pro*. Thus, e.g., the subject of a thetic or a presentational sentence cannot be directly taken up by *pro* (Frascarelli 2007, 705). The same holds for Familiar Topics in case there is a closer Aboutness-shift Topic available (Frascarelli 2007, 706; 709ff). A backgrounded subject, not

being an Aboutness-shift Topic, cannot be the local antecedent of *pro* (Frascarelli 2007, 708f). Therefore, the condition cannot be reduced to syntactic functions. Nor can it be reduced to definiteness, because an Aboutness-shift Topic in the form of a specific indefinite does qualify as antecedent of *pro* (Frascarelli 2007, 705). A *postverbal focused* DP takes the role of an Aboutness-shift Topic by default (unless there is an explicit Aboutness-shift Topic available), and thus can become the antecedent of *pro* (Frascarelli 2007, 726); a preverbal focused DP cannot by itself start a topic chain.

What makes this approach "multifactorial" is its syntactic implementation. Frascarelli (2007) assumes that *pro* agrees in ϕ-features with (the head of) a discourse-oriented position for Aboutness-shift Topics at the left edge of the sentence (Shift^0, leftmost in a split CP structure). Every sentence has a (potentially empty) Aboutness-shift Topic in this position, which can be coreferentially matched with the closest preceding Aboutness-shift Topic. *pro* is identified by this Agree relation, rather than by verbal agreement features.

3.3.4 Summary

In this section, we summarised and reviewed the most important general linguistic approaches to the phenomenon of referential null subjects. Four major threads of thought were distinguished. (i) Proponents of a *contextual licensing* view draw attention to the potential intervenors in the textual anaphor-antecedent relation: clause and sentence boundaries, textual unit breaks, intervening potential antecedents. Furthermore, they consider semantic properties like saliency, animacy, agentivity etc. which may forge or hinder this relation. (ii) *Grammar-oriented* research into null subjects has concentrated on the question what other grammatical properties may correlate with the setting of a pro drop parameter cross-linguistically. Proposals include the availability of subject expletives, subject extraction, and inversion; the richness of verbal agreement; and the internal morphology of pronouns. Critique of the alleged empirical correlations has let to much theoretical and technical refinement until very recently. (iii) Multifactorial approaches try to combine the grammar-oriented view with insights from the contextual licensing perspective, in order to explain the systematic aspects of pro drop vs. pronominalisation in actual texts. We will adhere to this strategy in the remainder and apply them to diachronic corpus data for Ru, Pol and Cz. To this end, the most important factors from the approaches discussed here have been combined into an annotation scheme which can form the basis for corpus queries and quantitative investigations. We will return to the details of this scheme after discussing what is known about null subjects in Slavonic languages *synchronically* in the next section.

3.4 Referential null subjects in Russian, Polish and Czech: Synchrony

In this section, we discuss the null subject properties of modern Ru, Pol and Cz, as a prerequisite to the diachronic analysis in 3.5. While Cz and Pol are fairly uncontroversial canonical null subject languages, there is ongoing disagreement in the literature about the status of Ru. Although our focus is diachronic, we hope to add some clarification to the opposing standpoints concerning modern Ru as well. We go through the presentation language by language, with contrastive or typological analyses being discussed where they become relevant.

3.4.1 Czech

3.4.1.1 Grammatical aspects of pro-drop

Modern Cz shows all the hallmarks of a canonical null subject language in the sense of the parametric approach to pro-drop (cf. also Adamec 1959; 1988; Uhlířová 1987; Kresin 1994; Lindseth 1998):

Cz verbal paradigms show no syncretism with respect to grammatical person, and very little gender and number syncretism: In present tense, the 3SG and (one variant of) the 3PL in class IV (*vědí/ví 'know-*3PL*', ví 'know-*3SG*'; prosí 'ask-for-*3SG/PL*'*) overlap in form; in past tense, there is gender/number syncretism between F.SG and N.PL forms, as well as between M.PL.INANIM and F.PL forms , in the standard variety.

Leaving a subject pronoun unrealised does not lead to a loss in grammaticality, as long as certain restrictions on coreference in context are respected (cf. 3.4.1.2). Coreferent subject pronouns are usually null in the standard language, realisation being the marked option. As in other canonical null subject languages, pronominal subjects in embedded adverbial clauses are understood as disjoint in reference from the matrix subject (on a parallel case in Italian see Roberts and Holmberg 2010, 7 and Frascarelli 2007):

(18) Co si *pro$_i$* myslí, když ona$_{j \neq i}$ jí$_{i/k \neq j}$ píše jako pětisetstránkovou
what REFL-DAT think-3SG when she her-DAT writes MP 500-pages-
omluvenku?
excuse-ACC.SG
'What does she$_i$ think when she$_{j \neq i}$ writes her$_i$ a 500 pages excuse?' (after ČNK)

For *ona* in (18), this indexing is the only possibility; the example could never express coreference of the embedded subject with the matrix subject.

Incidentally, much the same holds for complement clauses, cf. (3), repeated below:

(19) Rozeznal   *pro~i~* vysoké   třídílné   zrcadlo   a   došlo
      recognize-PT.M.SG   high-ACC.SG three-folded-ACC.SG mirror-ACC.SG and occurr-PT.N.SG

      mu~i~, že *pro~i~*/ on~j≠i~ leží   naznak   ve své   posteli
      him-DAT that   he-NOM lie-3SG backwards in self's-PRP.SG bed-PRP.SG
      'He recognized the high three-fold mirror and it occurred to him that he (= someone
      else) lay on his back in his own bed.'

Cz also meets the other conditions of the "pro-drop syndrome" mentioned in 3.3.2.1: It has
null expletive subjects (20-a), does not show *that*-t effects (20-b) – cf. also Meyer (2004) for
details –, and can freely invert subjects (20-c):

(20) a. Pršelo.
          rain-PT.N.SG
          '(It) rained.'

      b. Myslela   si: Co   očekávají, že   budu   chtít?
          think-PT.F.SG REFL-DAT what expect-3SG that AUX-FUT.1SG want
          'She thought: What do they expect that I want?'   (ČNK)

      c. Zavolalo   mnoho studentů.
          call-PT.N.SG many   student-GEN.PL
          'Many students called.'

The only fact which might cast doubt on the pro-drop status of Cz is the existence of
optional expletive-like elements. In chapter 4, we will see that these are not proper expletives
(following Lindseth and Franks 1996; Lindseth 1998, contrary to Kosta 1990). Diachronically,
they probably developed from attributive demonstrative adjectives. Thus, Cz can be regarded
a fully canonical NSL.

3.4.1.2 Information-structural and discourse restrictions

Turning to the conditions on null subjects in context, we find that in cases of topic continuity
across sentence boundaries, a null subject is obligatory in Cz; as Uhlířová (1987) puts it,
Cz overt subject pronominalization always expresses *discontinuity* [of the current discourse
topic]. Furthermore, null subjects are, as usual, excluded under focus (Adamec 1959, cited
after Kosta 1990) or when the subject forms a contrastive topic (Adamec 1988).

Kresin (1994, 76ff) names the following types of context which allow for null subjects in
modern Cz, but typically demand subject pronominalisation in Ru:

- trans-sentential coreference with an item which was just previously introduced (e.g.,
  by a presentative statement)

(21) Z   domu   vyběhl   pan kapelník Petrbok$_i$ v uniformě a
      from house-GEN.SG run-out-PT.M.SG Mr musician P.-NOM   in uniform-PRP and

      utíkal   na náměstí. Ø$_i$ Měl   bílé rukavice.
      make-off-PT.M.SG to square-ACC   have-PT.M.SG white gloves-ACC.PL

      'Mr. musician Petrbok came running out of the house in his uniform and made off

      to the square. He was wearing white gloves.'   (ČNK)

- trans-sentential coreference with a non-subject

- coreference (from inside a subordinate clause) with a main-clause non-subject (cf. also
  Koktová 1992):

(22) Zavolal   jsem   maminku, a   řek   jsem   jí$_i$,   aby   Ø$_i$ mně
      call-PT.M.SG AUX-1SG mum-ACC   and tell-PT.M.SG AUX-1SG her-DAT that-SBJ   me-DAT

      přinesla   ručníky.
      bring-PT.F.SG   towels-ACC.PL

      'I called mum and told her that she should bring me towels.'   (ČNK)

- coreference with a preceding (specific) indefinite:

(23) Na druhé straně   někdo$_i$ dlouho mlčel.   Pak Ø$_i$ řekl:
      on other side-PRP.SG someone-NOM long   be-silent-PT.M.SG then   said-PT.M.SG

      'On the other end of the line someone was silent for a long time. Then he said: '

                                                          (Kresin 1994, 79 (9))

- Following a direct speech event, a 3rd person speaker must be identified by pronominal
  or nominal reference in Ru, but is preferably a null subject in Cz (cf. Uhlířová 1987, 29)

- coreference with a general discourse topic, even if it is several sentence boundaries
  away and other topics intervene (cf. the Cz translation in (24-b) to the Ru original in
  (24-a)):

(24) a. Priexal   Ordžonikidze. [6 sentences] On   sidel   za
          come-PT.M.SG   O.-NOM   he-NOM sit-PT.M.SG behind

          gromadnym   pis'mennym stolom [Ru]
          huge-PRP.SG desk-PRP.SG

                                                          (Rybakov: Deti Arbata, cit. Kresin 1994)

      b. Přijel   Ordžonikidze. [6 sentences] ø/*On seděl   za
          come-PT.M.SG   O.-NOM   sit-PT.M.SG behind

          obrovským   psacím stolem [Cz]
          huge-PRP.SG desk-PRP.SG

          'Ordžonikidze arrived. [...] He sat behind a huge desk [...]'

                                                          (Kresin 1994, 97/98)

This last example best lends itself to an explanation via the hierarchy model (Fox 1987) introduced in section 3.3.1: As long as the overall topic of a whole text-structural unit – in this case, a paragraph – does not change, coreference can be expressed by a reduced form. Obviously, Cz *pro* thus does not depend on identification by a local Aboutness-shift Topic in exactly the same way as Italian does (Frascarelli 2007, 3.3.3): It seems to be sloppier about coreference with newly introduced items in presentational statements (21) and tolerates intervening Aboutness-shift topics, as long as the overall discourse topic is continued.

3.4.1.3 Overt subject pronouns

Cz has an especially rich system of demonstratives. On top of a demonstrative signalling proximity (Cz *tento*; Ru equivalent *ètot 'this (one here)'*) and one encoding distance (Cz *tamten*; Ru equivalent *tot 'that'*), it uses a third, locationally neutral, demonstrative *ten*. *Ten*, moreover, seems to carry a certain colloquial stylistic marking and a contrastive overtone. The difference in meaning between *on* and *ten* is paraphrased by Kresin (1994; cf. also Adamec 1983) as one between determining the *identity* of a referent (*on*) vs. a more "tokenizing" or "role-base" view (*ten*), which focuses on relevant properties of the referent or his "type membership". Cz *on*, in her analysis, instructs the addressee to "add the piece of information now being presented to a previously existing stock of knowledge of the referent" (p. 117) and therefore, presupposes knowledge of the referent beyond a single mention. *Ten*, on the other hand, links to an item, the relevant view of which, is "situationally bound role" is directly presented in the antecedent. It also retains a deictic component, which makes it appropriate for various "emotionalized" contexts (Kresin 1994, 106ff; cf. also critically Berger 1993). Inanimate coreference is generally expressed by *to* rather than *ono* (Kresin 1994, 109).

As far as differences among the three demonstratives are concerned, neutral *ten* marks coreference with items from the immediate or broader discourse and emphasizes their type membership. Proximal *tento* and its colloquial variant *tenhle* highlight the physical immediacy of the antecedent, and are favoured over *ten* when tokens of the same type are to be contrasted (Kresin 1994, 215ff). All these functions are covered by *ètot* in Ru. The deictic component of Cz *ten* is often weakened, rendering it semantically somewhat similar to a definite article (Adamec 1983). *Ten*-anaphora, including combinations with proper names, may reactivate a previously given or commonly known situation as a whole, rather than single out a specific referent:

(25) a. [Ne, vám   že to   budou   dávat s   nima,]   v
          no   you-DAT.PL that that-ACC AUX-FUT.3PL give with them-3PL.INS in

té   Vídni.
that-F.PRP.PL   Vienna-PRP
'[No, the three of you will have plays put on.] In that Vienna. right?'

(Kohout, cit. Kresin 1994, 234)

b.   Koupili jste   to   pivo?
buy-PT.PL   AUX-2PL   that   beer-N.SG.ACC
'Did you buy the/that beer?'   (ibid. 237)

To a somewhat greater extent than Ru *étot*, Cz *ten* is used with antecedents whose identity is resolved merely via associative links with known referents or situations. As a determiner of an NP modified by an attribute, it marks the selection of the intersection between the denotation of the noun and that of the attribute, as opposed to other subsets of the noun denotation, again much like an English definite article. However, Kresin (p. 266) maintains that *ten* differs from a definite article in at least two respects: (i) It always involves a deictic component, pointing to a *specific* set member; and (ii) it is realized only optionally.

To summarise, Cz is a clear pro-drop language, which, however also uses expletive-like elements with some unusual properties. Coreference relations are partly sloppier than in other pro-drop languages (Italian). Demonstratives form a rich system, distinguishing (at least) those marking mere coreference (*on*) and those focusing on further properties of the antecedent (*ten, tenhle, tento* etc.).

3.4.2 Russian

3.4.2.1 Grammatical aspects of pro-drop

As mentioned before, it is a matter of ongoing debate whether Ru should be classified as a canonical null subject language or not. The mere frequency of subject pronouns vs. null subjects in texts is not very conclusive: For the most part, *standard* Ru obligatorily realises pronominal subjects. Weiss (1993) stresses the point that *colloquial* Ru, however, allows for massive ellipsis – not only of argument expressions, but also, e.g., of verbs and complementisers. McShane (2009) claims that it is preferably 1/2P (rather than 3P) subjects which are elided in colloquial dialogues; but in principle, referential null subjects can occur in any person, gender, or number. According to Fehrmann and Junghanns (2008), argument ellipsis can be found completely regularly also in texts of the standard variety, cf. subject drop in (26-a,b) and object drop in (26-c). Note that (26-b,c), quoted from Fehrmann and Junghanns (2008), belong to a spoken register, but are definitely not substandard.

(26)   a.   Ivan   skazal,   čto _ pridet.
           I-NOM say-PT.M.SG that   come-3SG
           'Ivan said that he would come.'   (Fehrmann and Junghanns 2008, 193 (6))

       b.   _ Peredaju, čto   _ skazal.
            report-1SG what-ACC   say-PT.M.SG
            'I am only telling what he said.'   (Rybakov, cit. ibid. 193 (7))

       c.   [– Taz vynesu, togda ljagu, – otvetil Saša. – "I'll carry away the bowl, then I'll

            lie down, Sasha said.]

            – Ja   vynesu   _, – otec Vasilij podnjal   taz.
              I   carry-away-1SG   father V.   take-up-PT.M.SG bowl
            ' "I will carry (it) away". Father Vasilij took the bowl.'

                                                        (Rybakov, cit. ibid. 202 (18))

If the type of omission present in Ru affects all types of arguments alike, it might be a general strategy of contextual ellipsis or the grammatical property of "radical pro-drop" in the sense of Neeleman and Szendrői (2007). Neeleman and Szendrői (2007) themselves argue against a radical pro-drop analysis for Ru (see 3.3.2.6). Fehrmann and Junghanns (2008) opt for an analysis in terms of ellipsis, along the lines of Lindseth and Franks (1996) and Lindseth (1998). Franks and Lindseth had proposed three conditions for "real" pro-drop:

(27)   (i)   Only null pronominals are stylistically unmarked.
       (ii)   Only null pronouns can function as bound variables.
       (iii) Only null 3rd plural pronominal subjects can have arbitrary reference.

(27)(i) is obviously violated in Ru, since the standard even prescribes the use of pronominal, rather than null, subjects. (27)(ii) does not hold either, as we already saw in (2-c), repeated below:

(28)   Stepa   nakonec uznal   tjurmo   i   ponjal,   čto on
       S.-NOM finally recognize-PT.M.SG mirror-N.ACC.SG and understand-PT.M.SG that he
       ležit   navznič' u   sebja   na   krovati   [Ru]
       lie-3SG backwards at self-GEN on bed-PRP.SG
       'Styopa finally recognized the pier-glass and realized that <u>he</u> was lying on his back
       in his own bed.'

(27)(iii) is intended to ban overt 3PL pronouns from an arbitrary interpretation. According to Lindseth (1998), Ru violates this constraint, allowing an arbitrary reading in both cases of (29):

(29) Vo Francii _/ oni edjat ulitok.
      in France   they eat-3PL snails-ACC/GEN
      'In France they eat snails/snails are being eaten.'(Lindseth and Franks 1996, 202 (6c))

As Fehrmann and Junghanns (2008, 201) note, (29) may be acceptable in this reading for some speakers, but "it is not in line with normative grammar", which would prescribe a 3PL null subject here. So by the criteria in (27), Ru would be a non-NSL, and all factual argument omissions would have to be instances of contextually licensed ellipsis.

It is exactly the arbitrary null subjects – traditionally called *neopredelenno-ličnye* and *ob-obščenno-ličnye predloženija 'indefinite-personal / generalised-personal sentences'* – which led Müller (1988; 2006; 2007) to the assumption that Ru must be a pro-drop language. There is no way for (30) to have been derived from sentences with an overt subject pronoun by some process of contextual ellipsis, because the version with a pronoun would have a completely different – a coreferential – interpretation:

(30) a. Ob ètom *pro* mnogo govorjat
          about this –   much talk-3PL
          'There is much talk about this.'

      b. Ètu knigu *pro* pročityvaeš' za dva časa
          this book –   read-2sg   in two hours
          'This book can be read in two hours.'   (Müller 2007, 6f. (4))

Interestingly, the same observation is used as an argument *against* Ru being a pro-drop language by Lindseth and Franks (1996), who assume that the presence of a *pro* in (30) would obligatorily lead to a (wrong) referential interpretation, as in other cases of referential null subjects. Therefore, (30) could not contain *pro*, but only some other kind of null subject (a *pro*$_{arb}$, or possibly something like Mel'čuk's (1974) *Ø-ljudi 'people'*).

What (30) really shows, depends on whether we want to assume one *pro* which can end up either as referential or as indefinite/generalising, or several *pros*, each with a different interpretation. The system proposed by Holmberg (2005; 2010) (see 3.3.2.4) implements the idea that *pro* may be inserted without fixed referential properties (in a partial null subject language). If there is a local antecedent available, *pro* will turn out as coreferential, otherwise it will be interpreted as indefinite/generalising. Of course, one might also postulate a non-null subject language with a special indefinite *pro* in this system, losing, however, the further generalisations connected with partial null subject languages (see below).[5] Fehrmann and

---

[5] The third option in this system, a true pro-drop language with an additional indefinite *pro*, would be more problematic: The hallmark of canonical pro-drop languages, the [+D] feature in T^0, would exclude the indefinite *pro* (all else being equal).

Junghanns (2008, 191) propose exactly this – a non-pro-drop language with special pro$_{arb}$ items in its lexicon. A [+pro-drop] analysis has been adopted by Růžička (1986a), Kosta (1990) and Müller (1988; 2006; 2007).

Let us now turn to the other phenomena bundled with the "pro-drop syndrome": First, it is not completely clear whether Ru has "rich" or impoverished verbal inflection. Lindseth (1998) views Ru past tense markers as strongly syncretic with respect to person – they only distinguish number and gender. On her view, this "poor" agreement prevents Ru from being a pro-drop language. Baerman and Brown (2011) classify Ru person/number marking as non-syncretic, probably disregarding past tense completely. As we saw above in 3.3.2.3, Müller (2006; 2007) argues that the *l*-participle of the verb never contained any person features, nor does it today. In earlier stages of the language, person was marked on the auxiliary, which was, however, dropped during the diachronic development. So there would be no impoverishment, parallel to the Cz case.

Furthermore, concerning properties figuring prominently in generative theories, it is uncontroversial that Ru has null expletive subjects (31-a), and that free subject inversion is possible (31-b). Ru shows *that*-t effects with extraction from embedded clauses (Müller 1993; Meyer 2004) (31-c,d):

(31)   a.   Sverknulo   i   udarilo   nad   samym xolmom.
               flash-PT.N.SG and thunder-PT.N.SG above itself   hill-PRP.SG
               'There was a flash and a thunderclap right over the hill.'   (MiM, ParaSoL)

        b.   priexalo   mnogo studentov
               arrive-PT.N.SG many   student-GEN.PL
               'Many students arrived.'

        c.   *Kto$_i$ vy   xotite,   čtoby   t$_i$ vam   pomog?
               who-NOM you want-2PL that-COND   you-DAT help-PT.M.SG

        d.   Komu$_i$ vy   xotite,   čtoby   on   pomog   t$_i$?
               who-DAT you want-2PL that-COND he-NOM help-PT.M.SG
               'Whom did you want him to help?'

Unfortunately, these latter properties are not fully conclusive, since null expletives can also occur in non-null subject languages (e.g. in German), and there are uncontroversial pro-drop languages without *that*-t effects (Georgian, Finnish – cf. Gilligan 1987; Haider 2001, 285).

Roberts and Holmberg (2010, 11) include Ru among the so-called *partial* null subject languages, without providing arguments. The typical features of partial NSLs are the following (cf. Holmberg and Sheehan 2010, 126ff, see also 3.3.2.4):

(32)   a.   Null subject properties in some structural configurations, and non-null subject

properties in others

 b. i. In configurations in which a null subject is obligatory in a pro-drop language, it will be optional in a partial null subject language,

  ii. where it is optional in a canonical pro-drop language, it will be excluded in a partial null subject language.

 c. null generic pronouns

 d. identification of coreference by an antecedent in a higher clause

(32-a) has not yet been convincingly shown for Ru (see also above). Since this property is in some sense the defining feature of partial null subject languages, it may seem obvious that Ru cannot belong to this group. On the other hand, characteristics (32-b-d) are clearly met: Coreferent subjects in embedded adverbial clauses as in (1) (partly repeated below) must be null in Cz and Pol (33-b), but are only *optionally* null in Ru (33-a), cf. (32-b.i.) (see also McShane 2009):

(33) a. I ètogo sekretar' prestavit' sebe ne mog, xotja (on)
    and this-GEN secretary imagine SELF-DAT not can-PT.M.SG although he-NOM
    i xorošo znal prokuratora. [Ru]
    also well know-PT.M.SG procurator-ACC

  b. Ale nie mógł sobie tego wyobrazić, choć Ø nieźle
    but not can-PT.M.SG SELF-DAT that-GEN imagine although not-bad
    znał procuratora. [Pol]
    know-PT.M.SG procurator-ACC

  "And this the secretary was unable to imagine, though he knew the procurator well."

Generalisation (32-b.ii.), according to Holmberg (2010, 94), can be demonstrated by a variant of (33) in which the embedded subject refers to a person different from the matrix subject (cf. also his Italian example (16) above). Indeed, the prediction that Ru obligatorily uses an overt pronoun in this case, is borne out. The same point can be made when looking at the distribution of null subjects and subject pronouns in discourse (see below). (32-c) fits the picture, as discussed above (and in chapter 5). (32-d) is again evident from (33): Since the Ru 3P null subject cannot refer definitely to an individual or group all by itself, it requires a close antecedent (usually in the same clause), or it will ultimately support an arbitrary (/indefinite/generic) reading.

To sum up, evidence for Ru being a *partial* pro-drop language is convincing, except for the fact that no real "partiality" of null subjects in terms of a restriction to certain grammatical features on the verb etc. could be identified. We will stick to this hypothesis nevertheless in the remainder.

3.4.2.2 Information-structural and discourse restrictions

As far as pronominal vs. zero coreference marking is concerned, Grenoble (1998) departs from findings first established by Nichols (1984): Null anaphora are coreferent with the so-called *literary theme*, "the participant that a text or subtext is about" (Nichols 1984, 170), with "more formal literary styles tending toward overt marking and more colloquial styles tending toward zero marking" (Grenoble 1998, 189). Obligatory overt marking, according to Nichols and Grenoble, occurs (i) at the beginning of a new episode, (ii) after direct speech, and (iii) when there is a change in time reference (within an episode). Grenoble (1998, 190ff) adds observations based on spoken dialogues: In settings in which the interlocutors share much knowledge of referents, ellipsis is particularly frequent (p. 191); its occurrence also depends on the turn-taking structure, i.e., the current speaker is likely to be referred to by a zero anaphor and zero anaphors tend to occur turn-internally rather than across turn boundaries.

McShane (2009) scrutinises the conditions for object ellipsis in present-day Ru, Pol and Cz, as well as those for subject ellipsis in Ru vs. Pol from a largely theory-independent point of view. Her generalisations regarding subject ellipsis may be summarised as follows:

In Ru, the default is to realise the subject. Exceptions occur under the following conditions:

1. Subjects coreferent with another overt subject in coordinate structures up to a certain structural complexity, are elided.

2. In subordinate adverbial clauses, subjects corefent with an overt subject are optionally elided. Subjects of subjunctive complement clauses cannot be elided, but they cannot corefer with the matrix subject either (so-called *obviation effect*).

3. In a series of three or more related actions concerning the same subject, the non-first subjects are often elided. The same holds for textual elaborations, such as, e.g., re-statements of an event with details added; and also for true repetitions, i.e., structures lexically repeating the verb and the same overt or implicit arguments. In a similiar vein, subjects coreferring with a subject which is described extensively in the previous context are often elided.

4. Subjects with an antecedent in the same sentence, be it a coordination or subordination structure, may be elided. Across sentence boundaries, the main generalisation – alluded to by McShane as well – is that a deleted subject should continue a (chain of) subject(s) of thematically related sentences. Detailed description of an overt subject qualifies it as a new discourse topic, which may then become the antecedent of elided

subjects. Assuming that the above points are exhaustive, elided subjects in Ru can only have a subject antecedent or an *elided* object antecedent.

Kresin (1994, 17ff), following Nichols (1984), states that in hypotactic sentences, continuity in reference, grammatical role and discourse focus[6] tend to favour zero subjects in the embedded clause, while various types of discontinuity produce a tendency towards overt subject expression, with discontinuity of grammatical function being the strongest relevant factor. As an exception to this rule, objects introduced in the comment part of a clause may easily be referred to by a later anaphoric zero argument; cf.

(34) Na ulice   vstretil   Marusju$_i$. Ø$_i$ šla   s   koromysom   na pleče.
      on   street-PRP.SG met-PT.M.SG M.-ACC.SG   go-PT.F.SG with yoke-INS.SG on   shoulder
      Ø$_i$ ulybnulas'   emu.
      smile-PT.F.SG-REFL him-DAT

(Rybakov, cit. Kresin 1994, 33 (21))

On Kresin's view (p. 370), this exception is related to verbal semantics, which may create stronger expectations of continuous relevance in some cases (yet to be analysed in detail). At the same time, even the subject of a presentational verb-subject sentence cannot by continued across a sentence boundary by a zero form; it must be taken up by an overt pronoun (Nichols 1984, 176). Furthermore, nominal realization is typical at episode or chapter breaks, while pronominalization signals the change of sense unit and perspective typical for paragraph breaks. The latter tendencies can, however, easily be overridden in order to create effects of continuity.

3.4.2.3 Overt pronouns

Cross-sentential coreference can also be expressed by demonstrative pronouns. In this case, *ėtot* either expresses contrast against a set of alternatives or adds an evaluative, emotionalized overtone. *Tot*, the far more common demonstrative coreference marker, has no stylistic restrictions. It is the standard means of contrasting reference to a previous subject (taken up by a form of *on*) and a previous non-subject or subject of a subordinate clause (see Kresin 1994, 56ff for elaborations). Grenoble (1998, 198) demonstrates that in constructed test cases, non-nominative clausal topics also form the (only possible) antecedent of a *tot*-NP, so the decisive factor indeed seems to be grammatical rather than information-structural. When coreference expressed by means of *ėtot* and *tot* occurs in contrast, the former usually refers

---

[6]The term "focus" is used in the sense of the focus of attention of the hearer, not in its information-structural sense here.

to an item with higher thematicity in the discourse, i.e. closer to the center of attention (Kresin 1994, 65). The distinction of deictic proximity (*ėtot*) vs. distance (*tot*) carries over in different ways to prenominal adjectival (determiner) uses, as Kresin works out in detail. Importantly, the reference point from which proximity and distance are evaluated may vary – it may be an anchor in space or time, or, more abstractly, in personal perspective; it need not be identical with the deictic *origo* of the narrator, but may concern the point of view of a participant of the described action. The referent being pointed at need not be present in the discourse, but can be introduced via associative links. Both *ėtot* and *tot* fulfil the function of highlighting the type or qualitative role of the antecedent in the given context, rather than its mere identity (as it were the case in *pronominal* coreference marking). They may also introduce an new role description for an already specified item. Pronominal coreference usually marks continuing reference to the identity of an established item (p. 165) of high thematicity or salience. The prenominal demonstrative can clarify coreference to a less salient item in this case. According to Kresin (p. 169ff), certain inherently non-individuated and therefore non-salient referents such as, e.g., *tišina* as part of a setting or durative adverbials, do not lend themselves to pronominal coreference at all. Within larger units of discourse, prenominal demonstratives sometimes signal coreference to an item which had temporarily fallen out of the centre of concern; this may even involve the modification of proper names (otherwise highly individuated entities, favouring pronominal coreference).

Grenoble (1998, 72) challenges the view that the uses of *ėto(t)* and *to(t)* in discourse can be fully explained via (a generalization of) the feature [± proximal]. While this does hold in cases of direct juxtaposition of the two demonstratives, it leaves certain facts unexplained. E.g., as pointed out already by Padučeva (1981), *ėto* as the subject of a copula clause marks greater emotional distance from the point of view of the speaker, than would an anaphoric personal pronoun *on*.

3.4.3 Polish

That Pol behaves like a canonical null subject language with respect to referential pro drop, has been more or less taken for granted in the literature; the discussion has instead mostly centered around the remarkable properties of Pol *indefinite* null subjects. Academic grammars of the language form no exception here. Nilsson (1982) and Bartnicka et al. (2004), among others, note that in Pol, despite its generally being a null subject language, subject pronouns are often realised. McShane (2009), who provides the most detailed and empirically rich recent study of the phenomenon, based on corpus data and comparatively addressing Ru and Pol, arrives at the following generalisations:

1. In Pol, the default is the *non-realisation* of an overt pronominal subject. At the same time, subject ellipsis is generally avoided when it could lead to referential ambiguity. The subject must be realised if it bears company to a focusing particle (*i 'even', nie 'not', to 'it was/is', też 'too, as well'*), or structurally hosts an apposition, or in some cases of verb ellipsis.

2. Certain syntactic environments forge disjoint reference or coreference, and thus influence the choice of ellipsis vs. pronominalisation.

   a) A subject coreferent with a preceding nominal complement of a preposition, possessive nominal adjunct, or nominative participant of a copula construction, must be realised overtly.

   b) An elided subject in a subordinate clause is automatically interpreted as coreferential with the main clause subject (provided agreement between the two); disjoint reference has to be marked by overt subject realisation.

   c) Elided subjects in a sequence of (syndetically or asyndetically linked) main clauses are all interpreted as coreferential.

   d) A subject coreferential with a preceding rhematic (i.e., right-peripheral) subject may be *optionally* realised as a pronoun with no further semantic implications.

   e) Emphasis on the utterance as a whole facilitates the use of overt subject pronouns.

3. In *spoken* Pol, pronominal realisations covered by none of the above points occur (under circumstances yet to be clarified).

The conditions for the interpretation of *elided* subjects uncovered by McShane (cf. 2. above) do not differ too much between Ru and Pol. In both languages, null subjects are coreferent with realised subjects in the same clause or the immediately preceding main clause across a sentence boundary, unless a break of discourse units intervenes. Notably, coreference in Ru is more liberal, allowing also for object antecedents, where Pol only forges coreference between subjects (to the extent that McShane's analysis on this point is exhaustive; see 3.4.2.2 above for details on Ru).

What does differ between the two languages, however, is the interpretation of *realised* pronouns. According to McShane, Ru, but not Pol, can use a subject pronoun in cases of coreference across a subordinate clause boundary. This is, of course, exactly the issue illustrated in our introductory example (1). According to Holmberg (2010, 91ff), it constitutes a general difference between canonical null subject languages (like Pol and Cz) and partial null subject languages (Ru) – cf. the discussion in section 3.3.2.4.

Fehrmann and Junghanns (2008, 14ff) introduce a further fine distinction between Pol and Cz which had often gone unnoticed in the literature. They cite Górna (1976) with the observation that Pol allows a non-emphatic overt 3P pronoun in the preverbal subject position, as in

(35)   Parę dni   później on był   świadkiem   rozmowy.
            some day-GEN.PL later   he be-PT.M.SG witness-INS.SG conversation-GEN.SG
            'Some days later he did witness a conversation.'

                                  (Górna 1976, 212, cit. Fehrmann and Junghanns 2008, 215)

with focus on the verb (so-called Verum focus). This is impossible in standard Cz (although perfectly possible in Ru). A weak 3P pronoun may also be realised post-verbally in Pol (= *Parę dni później był on świadkiem rozmowy*) – again, as in Ru, but not as in Cz. A variant with a postverbal 1P pronoun, however, would be excluded in Pol and Cz (, although still possible in Ru).

(36)   *Parę dni   później byłem   ja świadkiem   rozmowy.
            some day-GEN.PL later   be-PT.M.1SG I witness-INS.SG conversation-GEN.SG
                                                    (ibid.)

The upshot is that Pol obviously has overt weak pronouns only for *part* of the person/ number paradigm, i.e., for 3P, but not for 1/2P. Standard Cz, on the other hand, has no overt weak pronouns, and Ru has them for the full paradigm. The situation in Pol thus conforms to a markedness scale evoked by Artstein (1999), if extended to weak subjects, as Fehrmann & Junghanns propose. Artstein claims that a first or second person null subject is less marked than a third person null subject; therefore, a language which has 3P null subjects will also have 1/2P null subjects. Transferring this statement to weak pronominal subjects, Pol is the case of a language providing overt weak pronouns only for part of the person / number paradigm, and thus it will have them for third person, not for first or second persons.

The fact that Pol uses weak subject pronouns, especially in the postverbal position, is noted also in Engel and Rytel-Kuc (1999, 64) (disregarding the person distinction); moreover, the authors mention that subject ellipsis is rather used in non-spontaneous registers, whereas weak subject pronouns are frequent in colloquial speech. On the other hand, a 3P pronominal subject in a complement clause cannot be coreferent with the matrix subject (Engel and Rytel-Kuc 1999, 65).

3.4.4 Summary

The preceding section, while far from being exhaustive, has introduced the main findings on null subjects in Ru, Pol and Cz present in the literature, and related them to general theoretic approaches where possible. In a nutshell,

Czech is a clearcut canonical null subject language: Overtly realised pronominal subjects signal a shift of the current Aboutness Topic (in Frascarelli's 2007 terms); null subjects signal topic continuity even across sentence boundaries and occasionally, even across intervening Aboutness Topics (i.e., under more liberal conditions than in Italian). Embedded clause subjects which are coreferent with the main clause subject must be null. Grammatically, all of the classically postulated correlations with expletives, extraction, and inversion hold. The only "irregularity" concerns the existence of apparent expletives with rather unusual properties (subject doubling and agreement, cf. chapter 4).

Russian is a borderline case, with a majority of researchers adopting an analysis as a non-pro-drop language with widespread contextually licensed ellipsis, but also important advocates of a pro-drop view. The existence at least of null subjects with an arbitrary or generic interpretation, which cannot possibly be derived by ellipsis, is put forward by proponents of the pro-drop analysis; the very same fact is taken as evidence by the non-pro-drop "school" that there cannot also be a *referential pro* in Ru. The contextual licensing conditions for coreferential null subjects across boundaries are not too unusual; but the distribution of overt coreferential pronouns differs drastically e.g. from the one in Cz: Overt pronouns do not forge disjoint reference except in subjunctive subordinate clauses. The evidence from *grammatical* correlates of pro-drop is not fully conclusive in Ru. Lately, an analysis as a *partial* pro-drop language has been proposed.

Polish has usually been viewed on a par with Czech, but recent research has uncovered a fine-grained distinction between the two: Polish has the *option* of using weak overt 3P pronouns in preverbal and postverbal position, which is impossible in Cz. It is often noted that Pol uses more weak subject pronouns than would be expected for a pro-drop language.

The textual, semantic and grammatical conditions deemed relevant for the analysis of null subjects are largely special cases of those proposed in general linguistic treatments. The analyses operate with concepts like textual units, sentence boundaries, and topic continuity; semantic properties like saliency or animacy are rarely invoked. A particularly important

area which does not always figure prominently in general linguistic approaches to null subjects is information structure. The factors present in the synchronic theories are mirrored by annotation categories in the corpus. Therefore, I annotated the corpus texts for verbal morphology, subtypes of null subjects, special syntactic constructions, and information structure. In the case of overt pronouns, properties of the antecedent (grammatical function, information-structural status) and the number and kind of intervening boundaries, are also coded.

As shown in the present section, modern Ru, Pol and Cz indeed differ considerably with respect to null subject properties or – even more clearly – in the licensing and interpretation of overt subject pronouns. We now turn to our central question, namely, how this present-day variation came about diachronically.

3.5 Referential null subjects in Russian, Czech and Polish: Diachrony

3.5.1 Russian

3.5.1.1 Previous studies

Null subjects are a long-standing topic of inquiry especially in the Ru linguistic tradition, and virtually every historical grammar or textbook contains some relevant remarks. The history of research on the matter has been broadly and diligently discussed before (cf., specifically, Kosta 1990 and the references cited there). For this reason, I will focus on the most detailed, empirically rich and influential treatments here, those by Borkovskij (1949, 1968, 1978) and, recently, by Zaliznjak (2008), referring the reader to the above mentioned work for a full research historical account. Since diachronic data (from Borkovskij's works) is also discussed in the generative analysis by Lindseth (1998) – an important analytical orientation point for the present thesis –, I will review her argument as well.

Borkovskij (1949, 90ff) analyzes null subjects (in his terminology: *opredelënno-ličnye predloženija* 'determinate-personal sentences') in a corpus of Old Ru *gramoty* (charters), i.e., documents belonging to an official, non-narrative register, which is less influenced by Church Slavonic. He finds that

1. elision of the subject pronoun in first and second person is a characteristics of the *gramoty*. An overt pronoun was unnecessary, because verbal agreement sufficed to mark the intended referent.

1SG/PL	Ø	pronoun under logical stress	pronoun for juxtaposition	unmotivated pronoun
before 15th c.	130 / 30	36 / 1	6 / 11	1 / 0
15th-16th c.	27 / 35	26 / 21	3 / 28	3 / 2

Table 3.2: 1. person subjects in *gramoty*, according to Borkovskij (1949)

2SG/PL	Ø	pronoun under logical stress	pronoun for juxtaposition	unmotivated pronoun
before 15th c.	25 / 16	3 / 1	1 / 0	0 / 0
15th-16th c.	5 / 34	3 / 4	0 / 2	2 / 17

Table 3.3: 2. person subjects in *gramoty*, according to Borkovskij (1949)

2. the use of subject pronouns was motivated by

   a) stylistic rules like specific stress/highlighting or juxtaposition (summarised under the heading of "logical stress" (*logičeskoe udarenie*))

   b) the overall prevalence of sentences structured into an overt subject and predicate (*dvusostavnye predloženija*)

   Both factors came into play at the same time and at some point drove the *odnosostavnye predloženija* out of the system in favour of the *dvusostavnye*. This change must have taken place around the 16th century and equally affected the whole territory (Borkovskij 1949, 105/106).

3. in 1SG/PL indicatives , 159 (68.8%) of 232 cases came without a pronoun, 62 realized a pronoun under "logical stress", 7 realized a pronoun for clarity of reference because of a change of the current topic, and only 4 (1.7%) realized a pronoun without any obvious motivation. In *gramoty* of the 15th-17th century, there was a rise in pronominalization from 24.8% to 54%. However, a large (not exactly given) number of examples in later documents consisted in the formula *ce азъ 'forsooth I'* (Borkovskij 1949, 94f).

4. In 3P, null subjects prevailed throughout. *Онъ 'he'* referring to a full noun phrase appears 33 times altogether, only 2 of which are demonstrative uses; 71 times the antecedent of *онъ* was itself a null argument. The majority of uses (87.1% = 27) of *онъ* were found in texts from the 15th-16th century, only 12.9% (= 4) come from 14th century documents. Regarding the grammatical function of the antecedents, 20 cases (64.5%) were subjects, and 10 objects (32.5%).

This study sets the scene for decades; later studies under the auspices of Borkovskij only add evidence from more texts, and repeat the overall generalisation under 2. above; they regrettably lack quantitative evidence, which the 1949 study provides at length. But the numbers in Borkovskij (1949) should be taken with a grain of salt, to say the least. E.g., the counts for 3rd person pronouns under 4. above are presented to suggest a rise of coreferential subject pronouns during the 15th-16th century. In fact, they are not very meaningful: The numbers have not been normalised for subcorpus size, and the size of the subcorpora do differ. But even if numbers were normalised, there is no reason to expect that each text contained an equal number of candidate sentences which open up a choice between a null or a pronominal subject. The 1/2ᴘ data is less conclusive than the author states: 2ᴘ pronouns do not show as clear a development as 1ᴘ pronouns do.

In his comparative East Slavonic diachronic syntax (Borkovskij 1968), the author applies a similar descriptive scheme to a larger corpus of Old Russian literature. He departs from a remark by Istrina (1923), according to which many instances of subject pronouns in Old Ru lack functional motivation (which would be an indication of a non-null subject language early on). By contrast, Lomtev (1956) reinforces Borkovskij's own (1949) conclusions that sentences *without* 1. and 2. person subject pronouns were prevalent in Old Russian and that the number of pronouns rose slowly from 13th to 15th century, but still then, had not become the norm.[7] Providing numerous examples but little quantitative evidence, Borkovskij (1968, 22–44) tries to make the point that the Old Russian norm up to the XV. century was the *odnosostavnoe predloženie,* i.e., a sentence pattern lacking a subject/predicate divide, and that the use of a pronoun could only be motivated by

(37) a. reference to a group including the author himself and the reader, in author's side remarks
      b. referential changes in the course of the narrative
      c. juxtaposition with a further referent in the discourse
      d. attributive qualifications like *(азъ) грѣшьныи 'I sinner',* often dictated by religious style
      e. emphasis on the subject
      f. attempt to soften a demand (in imperative clauses)

---

[7]It is somewhat bold to speak of an Old Ru "norm" for the 13th-15th century, given the problem of diglossia, the small amount of clearly East Slavonic texts and the lack of standardisation. Cf. Keipert (1999) for an overview of the variational situation in the Old Rus', and Rothe (2000) for a critique of the idea of an Old Ru literary language.

Cases (37-b), (37-c) and (37-e) are often accompanied by adversative particles and conjunctions like *но, же* or *и.* Borkovskij contradicts virtually all examples of "unmotivated" pronoun use adduced by Istrina, ascribing them "logical stress" on the pronoun, i.e., explicit or implicit juxtaposition or another form of emphasis. This state of affairs allegedly holds for the *žitija (vitae),* the chronicles, the *Slovo o polku Igorěvě,* the *Zadonščina,* and the *gramoty* on birch bark and on pergamen, and applies to both imperative and indicative or subjunctive clauses. Only the Novgorod chronicle (Istrina's major source), according to Borkovskij, might indicate an increase of "unmotivated" pronoun use since the 13th century.

For the 15th-17th century, Borkovskij (1968, 44) claims that there was no significant rise of unmotivated pronoun use either. However, in 15th-16th century *gramoty,* in the *Domostroj* and in the *Stoglav,* pronouns sometimes appear without motivation even with present tense forms, which are unambiguously marked for person and number. At the same time, null elements could still be used even in cases of juxtaposition of two subjects. Thus, Borkovskij (1968, 48) identifies about 300 *odnosostavnye predloženija* in the epistolar style of Ivan's IV. correspondence with Lord Kurbskij, several of them in juxtapositions, but only 60 cases of subject pronouns. As before, he argues that using a pronoun was dictated by the stylistic rules in (37). The concept of juxtaposition is stretched very far, covering virtually any occurrence of two different subjects in the same complex clause, cf.

(38) И   уже   не разумѣю,   ни чего   у насъ   хочещи.
      and already not understand-2sg NEG what-GEN at us-GEN want-2sg
      'And I do not understand any more, what you want from us (at ours).'

<div align="right">(Pis'ma Kurbskogo:114 – Borkovskij 1968, 49)</div>

Discussing instances of unmotivated subject pronouns in Ivan's correspondence, Borkovskij, in a footnote, alludes to a further potential reason for pronominalization, namely the lack of an auxiliary indicating person and number: *"Postanovka pervogo* ty *vyzvano otsutstviem vspomogatel'nogo glagola [...]"* – '*The insertion of the first* you *is caused by the lack of an auxiliary'* (Borkovskij 1968, 50, fn. 36). But of course, that could have been said of many of the earlier examples. In any event, it is a completely different kind of "motivation" than the ones mentioned above. It remains obscure why this factor should not govern the use of subject pronouns throughout, and how it might interact with the other, less structural factors.

Borkovskij attaches great importance to the writings of *Avvakum,* because of the author's ability to convey slight emotional differences in the written register. Summarizing the issue of *odnosostavnye predloženija* vs. subject pronominalization in the 17th century, Borkovskij

(1968, 53f.) states that again, the former were still the norm, although the latter did not form an exception and did not depend on stylistic motivation any more. In the spoken register, as represented in Heinrich Wilhelm Ludolf's *Grammatica Russica* from 1696 (Unbegaun, 1959), pronominalization was allegedly on the rise, with null subjects being still prevalent. Old Belorussian and Old Ukrainian documents, according to Borkovskij, do not show any different development. He concludes that, from the 11th to the late 17th century, the East Slavonic languages were canonical null subject languages, in which the use of a subject pronoun in 1st or 2nd person was merely governed by stylistic rules (*logičeskoe udarenie*). The issue of unambiguous person/number marking on the verb is left aside at this point.

Towards the end of the 17th century, subject pronominalization allegedly increased without stylistic motivation; presumably in the 18th century, the *odnosostavnye predloženija* turned into (*'prevratilis' v*', p. 77) *dvusostavnye* with a null subject. From then on, it was not the pronoun in an "unstructured" clause, but the null in a "structured" one that called for an explanation. Strangely, Borkovskij does not give a hint at when exactly this change might have occurred, nor any evidence for it from contemporary texts – presumably because the 18th century already falls beyond his realm of interest, the era of Old and Middle Ru (*drevnerusskij* and *starorusskij jazyk*).

3P null subjects are discussed separately, under the heading of subject omission or ellipsis in clauses which are structured into verb and subject (*nepolnye dvusostavnye predloženija* – Borkovskij 1968, 197–217). The author assumes a rise of subject *onъ* as a personal pronoun during the 15th and 16th century, leading to fewer subject ellipses. Subject ellipses typically occur if they are coreferential with a preceding subject or object, or their reference may easily be inferred from the context as a whole (Borkovskij 1968, 201); the latter option allegedly only holds for the *gramoty*.

Borkovskij (1978) essentially repeats the findings of Borkovskij (1968), adding some frequency information. E.g., the *Sinajskij paterik* contains 245 occurrences of 1/2P pronouns, but 1063 cases of null subjects (p. 192); the *Izbornik 1076* contains 35 instances of 1/2P pronouns vs. 401 null subjects, in indicative or subjunctive clauses. Following a rise in frequency of subject pronouns during the 15th-16th century, the 17th century is now regarded as the critical period of change. As part of the evidence, Borkovskij adduces colloquial data from a *Russkij rukopisnyj razgovornik XVII veka* (unfortunately inaccessible to me) and, as above, from Ludolf's Grammar. Across all tenses, the former shows a ratio of 1/2P *null* subjects to 1/2P *pronominal* subjects of 17.6%, but the latter a ratio of 54.4%. Looking only at the unambiguously person/number marked future and present tense, the former ratio amounts to 30.2%, and the latter to 123%. In other words, null subjects by far prevail in the *Russkij*

rukopisnyj razgovornik XVII veka, whereas they are excedeed in frequency by overt pronouns in Ludolf's Grammar. Without further comment, Borkovskij concludes that both sources simply show how null subject *clauses* were driven out of the language by the *dvusostavnye predloženija.*

As far as the third person is concerned, Borkovskij (1978) states that generally, null subjects occurred more often than subject prononuns in Old Ru (p. 312); only in everyday official / professional prose (*delovaja pis'mennost'*) could he observe a regular *coreferential* use of *онъ*. In spoken registers, including letters before the 18th century, and also in the writings of Avvakum, null subjects prevailed. According to Borkovskij (1978, 318), no substantial, only quantitative changes have taken place in the use of 3rd person null subjects up until today.

Throughout these three works, Borkovskij provides an invaluable wealth of examples, and discusses important factors which undoubtedly influenced the choice of pronominal vs. null subjects, as they actually continue to do today. The most interesting and important generalisation, on my view, is that 1/2ᴘ vs. 3ᴘ pronouns may have taken a fundamentally different development. However, there are also some points with which I would like to take issue, motivating the need for a fresh look at the phenomenon.

First, the use of quantitative data is less than fully convincing. The raw frequencies of subject pronouns are influenced by many contextual and grammatical factors, which may dictate their use also in an otherwise clearcut null subject language. Whether these factors come into play, is a matter of content and textual organization rather than of the grammatical system. Therefore, it is virtually useless to relate the frequency of null subjects to the frequency of *all* subject pronouns in a given text (as done in Borkovskij 1978), since an unknown amount of the latter could not ever have been dropped, for various reasons (e.g., because of being focused). Given restricted corpus sizes and (still) relatively small numbers of examples, there is little reason to assume that the amount of pronouns under conditions like contrastive focus or topic discontinuity remains constant across texts or that the respective differences simply level out statistically. Estimates of null vs. realised subject ratios derived in this manner are not comparable across subcorpora and therefore useless as evidence for a diachronic change. Instead, the frequency of null subjects may only be compared to those pronominal and nominal subjects which actually *could* have been dropped. If pragmatic and text structural factors are not being controlled for systematically, the provided frequencies could just as well be indicative of the content and rhetorical organization of the texts in which they occur. This shortcoming is all the more deplorable given the enormous amount of empirical work which went into the respective analyses and is now lost, apart from the examples cited in the above works.

Second, there are conceptual problems with the overall influence of analogy in *odnosostav-nye* vs. *dvusostavnye* sentences envisaged by Borkovskij. Generally speaking, it would be quite surprising if there was a diachronic tendency to lose "bisegmental" structures especially in Ru, which is so rich in "monosegmental" impersonals still today. Concerning referential null subjects, the argument that 1/2P subject pronouns were on the rise *because* the bisegmental sentences expelled the monosegmental ones, is virtually circular, since no substantial *independent* definitorial criteria of monosegmental vs. bisegmental sentences are provided. Furthermore, it is quite unclear where to look for the bisegmental model or source of the analogy. It could not have been sentences with a *realised* 1/2P subject, since those were very rare. But if it had been the 3P pronominal subjects, one would expect that the loss of null subjects should have started in 3P, contrary to fact (see below). And finally, if the "structured" sentences with *nominal* subjects had initiated the change by analogy, we would expect *pro* drop to be constantly on the decline, even cross-linguistically. However, the apparent inconsistency or tension in a system which has overtly realised nominal subjects, but null pronominal subjects, is present in any canonical null subject language, and need not lead to the loss of null subjects at all. I conclude that the assumption of an overall analogical extension of bisegmental clauses is a myth.

Third, the concept of *logičeskoe udarenie 'logical stress'* is never generally defined by Borkovskij, and notoriously unclear (cf. (38) above). It should be made explicit how *logičeskoe udarenie* is embedded into a linguistic theory (of any kind), into information structural and word order constraints, and how it could be recognized by independent researchers. For reasons of explicitness and especially for cross-linguistic comparability, it seems more helpful to stick to the definable subphenomena under this umbrella, such as, e.g., contrast, topic continuity, coreference, word order inversion, or the use of adversative particles. What 'logical stress' is intended to cover on top of those (and other) more or less well-defined phenomena, is some very unclear notion of *emphasis*. However, my suspicion is that the use of an actually superfluous subject pronoun as such raises a Gricean conversational implicature, causing the hearer to assign some kind of special importance to the utterance. In that case, the impression of emphasis would be just a consequence of the (grammatically unmotivated) use of a pronoun, not its explanation. If emphasis should have any explanatory value, something much more specific would have to be said; but no such elaboration has been provided by Borkovskij.

Finally, Borkovskij aims at a proper descriptive picture, rather than an explanatory account. His main goal is to identify reasons for the actual use of subject pronouns under the premise that Old Ru was a null subject language up until the end of the 17th century.

However, as argued in section 3.4 above, modern Ru does differ from clearcut null subject languages like Cz, and one would like to have an explicit analysis of what brought about these differences. Quite obviously, it is also desirable to understand how such an analysis fits into more general theories of language change.

Borkovskij (1949, 90f; 1968, 50; 1978) discusses (more or less explicitly), that non-emphatic subject pronouns increased in frequency as the perfect auxiliaries decreased – first in 3p, then in 1/2p. The loss of the present tense copula could have additionally supported this development. These are actually the predictions of the *verb impoverishment hypothesis*. Lomtev (1956) takes a different stance, namely, that the drop of the present tense copula was an epiphenomenon of the independent rise of pronoun realisation. Georgieva (1968, 24ff) points out that subjects in Old Ru are often elided in constellations which would not forge ellipsis in modern Ru any more:

(39) I vystupi mužь Volodimerъ₁ i uzrě ì pečenězinъ₂ i
     and enter-PT.M.SG man Vl.-NOM and see-AOR.3SG he-ACC P.-NOM and
     pro₂ posmějasja, pro₁ bě bo serednij tělomь
          laugh-AOR.3SG was namely middle body-INS.SG
     'And there entered the man Volodimer and the Pečeneg saw him and (he) began to laugh, because (he) was of little size.' (PVL, cit. Georgieva 1968)

On her view, the general reason for the increase in pronoun realisation and the decreasing use of the auxiliary/copula is to be found in the *recategorisation* of the 3p pronoun *онъ/- a/-o* from a demonstrative to a personal pronoun. We are thus left with three basic views, towards which later approaches need to orient.

Whether the actual empirical findings provide enough evidence for or against one of these views, is not immediately clear. On the one hand, there are very early instances of auxiliary drop with overt nominal or pronominal subjects, as Lindseth (1998, 59ff) points out (examples (40)-(41) cited after her):

(40) a. Glěb" knjaz' měril" more
        G.-NOM lord-NOM measure-LPT.M.SG sea-ACC
        'Lord Gleb measured the (distance to the) sea.' (Tmutarakan Stone, 1068)

     b. ja dalъ rukoju svojeju
        I give-LPT.M.SG hand-INS.SG self's-INS.SG
        'I gave with my own hand […]' (Mstislav's charter, 1130)

On the other hand, early instances of pronoun realisation *despite* the presence of the perfect auxiliary – figuring under the heading of *logical stress* in Borkovskij's work – seem to contradict the verb impoverishment hypothesis. Such an example is

(41) sego ty že esi xotelъ
      this-GEN you MP AUX-2SG woll-PT.M.SG
      'So this is what you wanted.'   (Lavr.Let., 1377)

The evidence for the alleged auxiliary drop thus looks less conclusive than first expected. Despite the unclarity induced by (40)-(41), Lindseth continues to argue that the changes in the morphological marking of past tenses, which did not take place in the same way in Pol and Cz, formed the reason for the differences in null subjects found today. The loss of the perfect auxiliary allegedly started in 3P and then spread to 1P and 2P (Černych 1954). Note, however, that the existence of an early overt 3P perfect auxiliary in East Slavonic is very doubtful, given that it is phonologically null already in the early Novgorod birchbark charters (except for one single example – cf. Zaliznjak 2008, 257). Lindseth (1998, 62) thinks that this state of affairs may be due to the fact that in 3P, non-emphatic subjects of various kinds – nouns, demonstrative and other pronouns – could compensate for a lack of overt person marking in many contexts. So person marking on the verb was far less often necessary in 3P than in 1/2P, where only the personal pronouns – which were used only emphatically at first – could serve this purpose. According to most authors (including Lindseth 1998), the use of weak personal pronouns was established slowly, as the use of perfect auxiliaries decreased. Opinions on the configuration of both overt subject pronoun *and* overt auxiliary differ somewhat: Ivanov (1990, 332-337), e.g., mentions it as typical for religious style.

Zaliznjak (2004, 170ff; 178ff) makes important generalisations about the use of subject pronomina in the birchbark charters, i.e., those remaining Old Ru sources which are probably least influenced by OCS:

(42) a. In the periphrastic past tense, auxiliaries are in complementary distribution with 1/2P subject pronouns. Generally speaking, 1/2P personal pronouns are usually elided. In this respect, the birchbark charters differ from the "high variety" of the chronicles. In 3rd person periphrastic past, neither subject pronouns not auxiliaries are realised.

      b. Aorists and Imperfects have been virtually non-existent since the earliest sources, except for some introductory phraseologisms (cf. also Remneva 2003).

      c. Subject pronouns are used in forms of address, under contrast or to mark a discontinuous topic; in later charters, they are possible everywhere (as in modern Ru). By contrast, subject pronouns appeared in higher frequency already earlier in the chronicles. In 12th-13th century, the DAT and ACC forms of the personal pronoun were still enclictic, but since the 14th century they have been almost

exclusively strong.

In the context of the diachronic development of the clitic past tense auxiliaries and the copula, Zaliznjak (2008, 241) extends these observations and develops a more detailed system of generalisations (interestingly, without the slightest hint at previous studies). The basic rule, again, is the omission of 1/2P subject pronouns in all tenses. At least in less stylistically elevated documents, 1/2P subject pronouns are in complementary distribution with the copula or the overt auxiliary in perfect tense. 3P pronouns are elided virtually everywhere except in combinations with the enclitic *že* or with *a* or *i*.[8] However, even in the Old Ru birch-bark charters, 1/2P subject pronouns *are* realised, either obligatorily or optionally, in several specific constellations. They are *obligatory*, according to Zaliznjak, in the following cases:

- emphasis

- contrast or juxtaposition with the subject of the preceding clause

- direct address (2P pronouns)

- after coordinating *i, ti* in the sense of *to, togda, v takom slučae 'so, thus, in that case'* following an explicitly marked topic (*čto kasaetsja 'as concerns'*)

- if the pronoun is a common element of two or more adjacent clauses

Optional realisation of the subject pronoun applies in the following cases:

- contrast or juxtaposition with the subject of the subsequent clause

- if the subject differs from the one of the preceding clause and is a topic (*tema*)

- if the predicate is ambiguous for person

- if the subject is syntactically modified

- if the subject stands at the beginning of a period of clauses and denotes the common topic of this period

The latter enumeration is intended as an open list. Diachronically, Zaliznjak assumes that the old rule of pronoun realisation which licensed them *only in the obligatory cases* mentioned

---

[8] Zaliznjak does not treat 3P pronouns in any detail. As we will discuss below, *že* marks a non-continuing topic (be it new or contrastive), and *a* and *i* in this constellation are focusing particles. Thus, the overt realisation of the pronoun is to be expected here.

above, gave way to a new rule which allowed subject pronouns to be realised virtually everywhere. The optional cases merely represent some relevant contexts for pronoun realisation which should be understood as early precursors of the new rule. Several single occurrences of pronoun realisation do not fit this pattern already in Old Ru. Zaliznjak remains unspecific about when and how the change from the old rule to the new rule took place, but speaks of an evolutionary process throughout the 11th to 17th century (p. 246).

The earliest Old Ru documents partly show an older, more literary pattern, namely, the consequent realisation of the copula/auxiliary, even when the subject pronoun is also present, as e.g. in the above "obligatory" contexts. The same holds for the OCS Codex Marianus (p. 247). Zaliznjak presents counts of three types of overt patterns of subject pronoun and copula/auxiliary in his corpus: (i) the model *dalъ esi 'give-LPT.M.SG AUX.2SG'* ([-subject, +aux]), (ii) the model *ty dalъ 'you give-LPT.M.SG'* ([+subject, -aux]), and (iii) the model *ty dalъ esi 'you give-LPT.M.SG AUX.2SG'* ([+subject, +aux]). The amount of occurrences of model (i) and (iii) *in "optional" contexts*, in relation to the frequency of all three models in these contexts yields a *'coefficient of copula/auxiliary preservation' (koèfficient soxranenija svjazok)*, which is then compared across several important texts. Thus, e.g., the First Novgorod Chronicle (11th-13th century), which contains 21 x model *dalъ esi*, 2 x model *ty dalъ* and 3 x model *ty dalъ esi*, would have a coefficient of 92% (= sum of model *ty dalъ esi* and *dalъ esi* / sum of all models = 24/26). The life of Avvakum (17th century), on the other hand, has a coefficient of only 1%. Zaliznjak's (2008, 248) numbers suggest a smooth linear decline with a somewhat steeper fall of the coefficient during the 16th century.

As far as 3P is concerned, the birchbark charters virtually contain no perfect auxiliaries, independently of subject realisation. In the other documents, availability of the perfect 3P auxiliary depends on style or register. The elevated literary texts – beginning with the OCS Codex Marianus, the Life of Methodius, and continuing through later strongly Church Slavonic ones – retain the auxiliary. The chronicles and later also Avvakum take an intermediate position: Their more elevated or more strongly Church Slavonic passages favour the overt realisation of the perfect auxiliary, whereas the rest leaves it mostly unrealised (Zaliznjak 2008, 260f). This is independent of whether a pronominal subject is realised or not.

Similar counts are undertaken separately for copulae with a predicative adjective or noun. Zaliznjak's conclusion is that the auxiliary system did not change much between the 11th and 15th century, whereas a swift change towards the modern situation took place during the second half of the 16th century. In 17th century, the situation was already close to the modern one. The loss of the copula took a similar path (happening first in the same contexts), but was accomplished earlier than that of the perfect auxiliary. In elevated literary style (i.e.,

under strong Church Slavonic influence), copulae and auxiliaries could still be realised well after the 16th century.

Zaliznjak's (2008) work constitutes an important model for our investigation. Since he is primarily interested in the loss of the auxiliary/copula, his study is partly complementary to the present one. However, he puts down a catalogue of criteria for the realisation of subject pronouns which is very relevant in the present context. The underpinning with quantitative data is another great improvement over earlier analyses. Nevertheless, Zaliznjak – deliberately, so it seems – completely refrains from any theoretical analysis, and does not link his observations to any of the models discussed in the literature. His perspective obviously differs from Borkovskij's (1949; 1968; 1978) and Lindseth's (1998), and is closer to Lomtev's (1956) views: It is not auxiliary/copula drop which leads to the rise of pronominalisation, but rather the other way round – changes in pronominalisation motivate the loss of an overt auxiliary/copula. Some open ends concern the exact status of the "optional" contexts mentioned above: Zaliznjak suggests that this subclass simply opened up diachronically to include just any context today (except, presumably, the "obligatory" contexts, which still demand a pronoun). But it seems that cases of topic shift and of syntactic modification actually *require* overt pronouns in modern Ru (cf. also Nichols 1984; Kibrik 1996), i.e., they moved from "optional" to "obligatory". After text junctures (pericopes), explicit mentioning of the current topic is at least strongly preferred. So what we actually seem to face is a repartitioning of the relevant contexts rather than a mere extension of optionality.

3.5.1.2 Hypotheses and open issues

Many factors influence the raw frequencies of pro drop, subject pronouns, and other realisations of subjects; only a selection of these is relevant for our analysis of grammatical change in this domain. Our task is to operationalise the hypotheses in such a way as to blank out those differences between the individual texts which may dictate the use of a pronoun or of a null subject even in a canonical null subject language. E.g., the fact that a chronicle uses more past tense forms or less existential sentences than a text which contains philosophical reflections, is not interesting as such for our investigation. It is not even relevant whether one author uses more 1P pronouns, in absolute terms, just because he happens to refer more often to himself than others do; it would of course be wrong to deduce a diachronic rise of 1SG pronominalisation on this basis. Instead, we need to focus on those configurations, in which there is a *choice* between the means of expression, and see how subject realisation in these configurations changes over time. It is instrumental that we compare the *relative* frequencies of different means of subject realisation, or use statistical methods which do so

implicitly.

Among the factors to be considered, the most delicate and problematic are probably the *information structural* ones. This domain is less clearly defined than e.g. the usual grammatical categories. An extreme case in this connection is constituted by Borkovskij's (1968) claim that subject pronouns only occur under "logical stress" (*logičeskoe udarenie*) throughout Old Ru up until the late 17th century – which is based on a maximally unclear notion and thus becomes unfalsifiable (see above). However, several objective criteria for the information structural status of referents can indeed be pinned down also in historical texts:

(i) Particles like enclitic *že* and adversative *a* signal contrast. Only in exceptional cases does the scope of contrast affect something else than the constituent preceding *že*; cf. the following early instance of *že* indicating disappointment of an expectation linked to the preceding proposition:

(43)   [Почто еси здѣ положилъ его? – 'Why did you bury him here?']

Яко   азъ   старѣйши   его   есмь,   ты   же   положилъ   его   на   моемь
though I   older   he-GEN be-1SG you MP   bury-LPT.M.SG him in   my
мѣсте.
place
'Although I am older than him, you still buried him in my place.'

(KievPat, RRuDi)

(ii) Sentence-final subjects of transitive verbs are minimally focused.

(iii) Verb preposing in declaratives signals so-called thetic structures without an overt topic (or with a situational grounding as topic; for details with respect to modern Slavonic languages, cf. Junghanns & Zybatow 1995; Zybatow & Junghanns 1998; Junghanns 2002) – cf. the first clause of (39) above for illustration.

It cannot be absolutely taken for granted that *pro* only had the status of a weak, coreferent background element throughout the history of Ru. Highly salient entities, like e.g. *Jesus* in the OCS New Testament texts, can be referred to by a null subject, which may even serve as a Shifting Topic (H. M. Eckhoff, p.c.; Eckhoff and Meyer 2011). So it is necessary to browse the null subjects in our text samples for such exceptional cases. It turns out that referential null subjects (*pro*) in 3P in our samples are almost invariably anaphoric, non-contrasted, non-emphasized background elements, as expected. In the vast majority of cases, they referred to the topic of the preceding sentence, i.e., signalled topic continuity. So they had to be compared to the class of anaphoric, non-contrasted, non-emphasized background subject pronouns. These could be mere background elements or Familiar Topics, continuing an

immediately preceding Topic.[9] The indicator of a diachronic development relevant for our purposes can then be defined as the relative frequency of *pro* vs. (various types of) pronouns with the same feature values, evaluated for different time spans. This indicator should then be related to other linguistic features, such as auxiliary realisation, cliticisation, frequency of indefinite impersonals and the like, to work towards an explanatory account. The whole procedure may appear pedantic; but it can easily be seen that a simpler approach would be error-prone: If we e.g. just looked at the relative amount of null subjects in *all* present and past tense clauses – rather than only those with a "weak" subject –, we would find a number of 20% for Nestor's chronicle (11th c.), 27% for A. Nikitin's *Choždenie za tri morja* (15th c.), and 29% for Avvakum's *Žitie* (~1670) – i.e., an apparent development *towards* a null subject language, contrary to fact.

A clarification is in order regarding the relation of grammatical person to the information structural and referential properties addressed in the remainder. It is obvious that a 1/2P pronoun always comes with deictic, text-external reference. It also represents easily accessible and "given" information, since the discourse participants can be retrieved at any time. However, this does not mean that it must be discourse-old and cannot enter any other information structural relations. A 1/2P pronoun may be used contrastively or introduce a shift of the current Aboutness Topic. I.e., although being familiar, it need not be a "Familar Topic" in a more technical sense.

3.5.1.3 Corpus investigation

Previous studies of the diachrony of null subjects in Ru (see above) have established (i) that weak pronominal subjects occurred more often in genres which were less influenced by OCS, i.e., rather in the *gramoty* than, e.g., in homilies or sermons. However, they do occur in the

---

[9]In a pilot study (Meyer 2009b), *nominal* realisations were also taken into account for test purposes. The argument for including them is that they actually just form yet another referential device on a par with pronouns and *pro*. If a language preferred the repetition of nouns over using a pronoun (as, e.g., Belhare – Bickel 2003; Stoll and Bickel 2009), and we just compared pronouns to *pro*, the apparent relative amount of pro drop would become misleadingly high. There are indeed examples of coreferential, repeated NPs, especially in the older chronicles (cf. (i)), but also e.g. in the tale of Frol Skobeev (1720).

(i)   [И прищедшю ему к Судомири рѣцѣ, и Ярославъ ис Кыева въ 7 день постиже и ту. – 'And when he had come to Sudomir river, Jaroslav from Kiev also reached it within 7 days.']

И   побѣди   Ярославъ   Брячислава.
and   defeat-AOR.3SG   Ja.-NOM   B.-ACC
'and Jaroslav" defeated Brjačislav".'   (Lavr.Let., RRuDi)

However, it turned out that for four larger Old Ru text samples from different centuries, the number of nouns as continuing topics was very low and did not influence the results. For the purposes of the present study, we therefore only include pronouns and *pro* into the picture.

chronicles, which are, generally speaking, written in a high variety of Old Ru. (ii) The 16th-17th century counts as a critical period, during which the strongest changes in null subject properties obtained. The present investigation departs from these claims and tries to verify or qualify them on the basis of an electronic corpus. Two major sub-studies are being conducted: (i) A diachronic analysis of samples from Old Ru narrative texts, and (ii) a study of null subjects in various genres of 16th/17th century Ru. Concomitantly, null subject properties as represented in single texts, are being discussed. Over 7000 clause units, sampled in 2-6 random blocks per text, from 16 representative corpus texts (not counting single *charters* and *Vesti-kuranty* as separate texts), were annotated and analysed in detail for (i) and (ii).

Pronominal subjects were annotated according to the following algorithm, applying to all nominative forms (in all orthographical variants) of *ja, ty,* and *onъ*.[10]

1. Mark up all forms preceding an adversative particle (*že*) as instances of a Shifting Topic.

2. Mark up all forms following an adversative connective (*a, no*) as candidates for Contrast or a Shifting Topic.

3. For all remaining forms,

    a) mark up special syntactic constructions which might demand an overt realisation of their head: appositions, relative clauses, correlatives etc.

    b) For the syntactically unmarked forms,

        i. mark up immediately preceding pericopes and clause boundaries

        ii. mark up explicit contrast with another referent in the context

        iii. mark up the most plausible information-structural function: (Non-Topic-)Background, Minimal Focus, Part of Focus Domain, Contrastive Focus, Contrastive Topic, Shifting Topic, Familiar Topic

4. For all 3ᴘ forms,

    a) mark up all pericopes and clause boundaries intervening between the form and its antecedent

    b) mark up the grammatical function of the antecedent

    c) mark up the most probable information-structural function of the antecedent

---

[10] We ignore demonstrative pronouns (notably *sь* and *tъ*) here for reasons of space and coherence. They demand an investigation of their own, in the context of other demonstrative means of reference; furthermore, they have already been scrutinised elsewhere (see Berger 1993 and references therein).

Steps 1 and 2 can be done automatically, whereas the remaining steps require careful inspection "by hand". Clause boundaries, text junctures, and grammatical functions might be annotated independently by a sentence splitter, tagger and/or parser, which, however, are subject to a considerable error rate, especially in older texts. Step 4 (c) is only done implicitly – by marking up the pronoun as a "Familiar Topic" if the antecedent was itself a Topic, and as a "Shifting Topic" otherwise. In the long run, it would be more useful to deduce these values from the annotation of the pronoun as Topic and the fact that another Topic intervenes between the pronoun and its antecedent. But that would require more annotation than could have been achieved for the purposes of this study. Under a full information-structural annotation, my annotation could easily be re-used to derive the information-structural function of the antecedent. In the remainder, I will discuss the state of the null subject property in samples from selected narrative texts, and then return to the issue of verb impoverishment.

The distinction between null vs. pronominal subjects in 1/2ᴘ on the one hand and in 3ᴘ on the other, which was evident from the birchbark charters, carried over also to the narrative text samples. These only yielded an interpretable picture when evaluated separately by grammatical persons, especially in the later periods.

3.5.1.3.1 Initial situation

Old Church Slavonic   The status of OCS as a canonical null subject language is beyond doubt, as has been repeatedly noted in the literature (Večerka 1996 and references therein). By common origin or translatory influence OCS follows Greek extremely closely in this respect. In a quantitative investigation based on the deeply annotated Codex Marianus of the PROIEL corpus (http://www.hf.uio.no/ifikk/english/research/projects/proiel/) and its Greek counterpart, Eckhoff and Meyer (2011) found that non-salient (not recently mentioned or generally known) referents, as well as non-topics or too distant antecedents usually demanded an overt pronoun. OCS was more restrictive in this respect than Greek, but also allowed for some optionality of 3ᴘ pronouns e.g. after pericopes or when referring to generally salient, but textually distant entities. Accordingly, differences between OCS and Greek, in the vast majority of cases, went in the direction of OCS using more (weak) subject pronouns. The upshot is that OCS was clearly a null subject language, but also optionally used subject pronouns to a somewhat greater extent than the closely related "paradigm null subject language" Greek.

Turning now to the oldest East Slavonic sources, I first evaluated the Novgorod birchbark documents published by Zaliznjak (2004; http://gramoty.ru).

Novgorod Birchbark charters   It is cumbersome to speculate about anaphoric linkes here, because the documents are for the most part too short for that. In a first step, I evaluated all LPT forms (= 337 relevant cases[11]) and recorded the patterns of realisation of auxiliaries and subject pronouns. <mark>Non-periphrastic forms prevailed over forms with an auxiliary (= the former perfect tense) since the earliest texts.</mark> This tendency gets continously stronger towards the 15th century – from 52% (58:53 cases) in 12th century up to 75% (39:13 cases) in 15th century. There are clear differences in the use of the auxiliary depending on grammatical person: In 3P, virtually no auxiliaries are used (4 cases with an auxiliary vs. 150 without), while in 1/2P, about 75% of the relevant cases contain an auxiliary (137 vs. 46). In 1/2P, there is an obvious correlation with subject pronominalisation: If an auxiliary is realised, then a pronominal subject rarely gets realised (7% of the relevant cases – 9:128). If, on the other hand, a bare LPT is used, the agreement features are usually displayed on a subject pronoun (in 64% of the cases – 28:16). Of the [-auxiliary, -subject] group of cases, only 3 are really pertinent, for various reasons; cf. (44-a) for illustration; examples like (44-b) represent a special case, a long, partly elliptical, formulaic list of LPTs:

(44)   a.   да пришли   · сороцицю   · сороцицѣ за|былє   ·
           mp send-IMP.2SG   shirt-ACC.SG   shirt-ACC.PL   forgot-LPT.M.SG
           'Please send me a shirt. Shirts, I forgot.'   [Gram 43]

       b.   взѧлъ   ѥсємь оу шюги двѣ коби   рожи   и · оу
           took-LPT.M.SG AUX.2SG from Š.-GEN two baskets-ACC.PL rye-GEN.SG and   from
           мики|тки на оузи взѧлє   коробию   рожи   · на ксоно|ви ·
           M.-GEN   of U.-PRP took-LPT.M.SG basket-ACC.SG rye-GEN.SG   at K.-PRP
           взѧлъ   · потори   коробии   рожи   · […]
           took-LPT.M.SG   one-and-a-half basket-GEN.PL rye-GEN.SG
           'From Š., I took two baskets of rye and from M. of U. one basket. At K., I took
           one and a half baskets of rye […]'   [Gram 689]

In 3P, referential null subjects are very frequent despite the complete lack of overt agreement marking (cf. (45) for illustration; 91% of 60 instances). In most cases, an antecedent is located within the same sentence. The amount of pronominally or nominally realised subjects is very slowly on the rise: from 54% (25:51 cases) in 12th century , to 63% (19:11) in 15th century. <mark>The personal pronoun *on (он, ѡнъ)* was used extremely rarely, e.g. to mark a change of topic</mark>:

---

[11]"Relevant" cases were those which could be categorised clearly – taking into account also the available translations into modern Ru – showing no damage in places where an auxiliary or a subject pronoun was to be expected. Copulae were systematically distinguished from auxiliaries; the counts below only hold for the latter.

(45) [язъ юму ѿвѣчалъ   не реклъ   ми   есифъ варити перевары
      I   him   answer-LPT.M.SG not say-LPT.M.SG me-DAT E.-NOM cook-INF liquor-GEN

      ни на кого] ı   онъ прислалъ   къ федосы вари   ты пивъ
      NEG for whom and he send-LPT.M.SG to F.-DAT cook-IMP.2SG you beer

      '[I answered him: Esif" did not order me to brew liquor for anybody] and he sent to
      Fedosja "Brew the beer" '   [Gram 3]

In a second step, I analysed all occurrences of subject pronouns (= 145 cases). As far as
present tense copula constructions are concerned, no clearcut tendency to drop the copula,
nor any diachronic development to this end could be found (11 zero copulae, 7 realised ones).
The overall amount of subject pronouns in 1/2P in relation to the number of tokens per cen-
tury shows a slight increase (11th century: 0,24%; 12th: 0,51%; 13th: 0,99%; 14th: 0,97%; 15th:
1,22%). But, as discussed above, this type of counts can be influenced by many factors and
does not provide convincing evidence for a development towards subject pronominalisation.
There are only three instances of 3P subject pronouns, two of which are unclear and might
rather constitute expletives. Oblique forms of the pronoun are, by contrast, widespread from
the beginning. The birchbark charters thus provide too little evidence bearing on the devel-
opment *onъ*.

How can these findings be interpreted, and how do they tie in with Zaliznjak's (2008) gen-
eralisations? The almost perfect complementary distribution of subject pronouns and auxil-
iaries in 1/2P perfect forms may be taken as evidence in favour of the verb impoverishment
approach, but it remains indifferent towards the actual direction of influence: Did changes
in the pronominal system make person/numer marking by an auxiliary redundant, or was
it changes in verbal morphology which influenced subject pro drop? The former would be
closer to Zaliznjak's (2008) description, the latter rather in the spirit of the verb impoverish-
ment hypothesis. Some examples do not fit this clearcut picture, i.e., realise no agreement
marking at all, or contain a subject pronoun in presence of an auxiliary. We suggested above
that the former group can be neglected, because most of these cases may be explained away
as formulaic omissions or contextual ellipses.

As long as the subject/auxiliary system in 1/2P is so consistent as in the birchbark docu-
ments, 3P information can be reliably derived from the lack of an overt person marker, i.e., a
null morpheme in the auxiliary paradigm. The system is generally reminiscent of present-day
Common Cz (*obecná čeština*), which also has the option to use 1/2P pronouns and concomi-
tantly drop the perfect auxiliary, where standard Cz applies *pro* drop with an overt auxiliary;
3P lacks an overt marker on the auxiliary in both varieties (as in did in the Old Ru of the
birchbark charters). Given the large amount of referential null subjects in 3P, which do not
seem to require a close antecedent, the *pro* drop status of the birchbark charters is beyond

any doubt.

It must further be noted that the birchbark charters, although spanning a time period of more than four centuries, only provide very little evidence for a diachronic development from null to pronominal subjects. If we wish to draw diachronic conclusions despite the low amount of examples, a certain tendency towards auxiliary drop arises between 14th and 15th century, and a slight overall increase in subject pronominalisation in 1/2P can be recognised.

3.5.1.3.2 Samples of narrative texts

Early hagiography: The Kiev Paterikon   The Kiev Paterikon (*Kievskij Paterik*) is a collection of uplifting and instructional narrations from the lives of saints, compiled in the Kiev Caves monastery in early 13th century. Its language is strongly influenced by Church Slavonic, AOR is the predominant preterite form, and pro-drop dominates over subject pronominalisation in all tenses.

	Aorist	Perfect	Past	Present
1/2P	21 (=100%)	17 (=100%)	1 (=50%)	82 (=98.8%)
3P	47 (=100%)	2 (=100%)	2 (=100%)	12 (=100%)

Table 3.4: Amount of pro-drop in the Kiev Paterikon sample (n = 602)

The percentages in the table represent the amounts of null subjects in relation to all relevant comparable subjects of the given category (i.e., coreferential background material or Familiar Topics). Thus, e.g., 82 (= 98.8%) of the relevant 1/2P subjects with a predicate in present tense were *pro*. Under a verb impoverishment approach, we would expect the amount of null subjects in the "Past" (= bare LPT) column to be minimal, or at least a lot lower than that in the "Aorist" or "Present" column. The Kiev Paterikon sample contains so few instances of bare LPT that it can barely be regarded as informative in this respect. Note in passing that in 3P, Perfect (i.e., periphrastic preterite with an explicit auxiliary) was possible, an option which we find very rarely in Old Ru texts. As far as weak subject pronouns are concerned, these are lacking entirely in our sample in 3P, and occur rarely in 1/2P – cf. (46) for illustration.

(46)   Повѣждь   мнѣ,   рабу своему,   язву   смертную, да аще азъ тя
      display-IMP.2SG me-DAT slave self-POSS.DAT.SG wound deadly-ACC.SG so if I you
      не изоврачюю, да будеть   глава   моа   за главу   твою   и
      not cure-1SG   so be-FUT.3SG head-NOM my-NOM for head-ACC your-ACC and
      душа   моа   за душу   твою!
      soul-NOM my-NOM for soul-ACC your-ACC

'Show me, your slave, your deadly wound, and if I do not cure you, then my head will be for your head and my soul for your soul.' (KievPat, RRuDi)

In the full text of the *Kievskij Paterik* (rather than only in the sample), only 3 of the 59 occurring 3P subject pronouns could be classified as Familiar Topics, i.e., as truly weak pronouns; the vast majority (53) were Shifting Topics, usually marked by *že*. This lack of "weak" uses essentially carries over to 1/2P pronouns, although those also occur without contrast or emphasis at the beginning of utterances, i.e., after a pericope:

(47) Она   же   рече: «Азъ   тя   искуплю,   и   славна   сътворю   тя,
      she   MP   say-AOR.3SG   I-NOM   you-ACC   liberate-1SG   and   famous-ACC   make-1SG   you
      'And she said: I will liberate you and make you famous,' (KievPat, RRuDi)

Chronicles The *Novgorod Primary Chronicle* (*Novgorodskaja pervaja letopis' staršego izvoda*, or *Sinodal'nyj spisok*, begun early 11th cent., earliest redaction 1260–1330) – the oldest remaining Old Ru chronicle – shows a similarly clear picture: There are virtually no weak subject pronouns in our sample, except for single instances in 1/2P, such as (48):

	Aorist	Perfect	Past	Present
1/2P	8 (=100%)	1 (=100%)	–	8 (=88.89%)
3P	165 (=100%)	–	2 (=100%)	7 (=100%)

Table 3.5: Amount of pro-drop in the Novgorod Primary Chronicle sample (n = 613)

(48) бояре   же […] не   даша   ему   напустити Фрягъ,   рекуче:
      boyars-NOM   MP   not   give-AOR.3PL   him-DAT   let-in   F.-GEN.PL   say-GER.PRS
      «мы   с   тобою   есмь».
      we   with   you-INS   be-1PL
      'But the boyars […] did not allow him to let in the Italians, saying: We are with you.'
                                                                                  (NovgStar, RRuDi)

The clarity of the result seems to contradict Istrina (1923) (who is criticised sharply by Borkovskij, see above), when she claims that the Novgorod Primary Chronicle (Synodal redaction) contains many functionally unmotivated uses of subject pronouns. Borkovskij's opinion was that all of these belong under the umbrella of "logical stress". Therefore, I reanalysed the uses of 3P personal pronouns in the full text of the chronicle, rather than only in the 600 clauses sample, according to many possibly relevant parameters, as detailed in 3.5.1.3 above. The distribution across information-structural categories is indicated in table 3.6:

	ConTop	ConFoc	FamTop	MinInfoFoc	InfoFoc	OpFoc	Bg	ShiftTop	Sum
3P	4	0	2	0	0	0	0	37	43

Table 3.6: Subject pronouns in the Novgorod Primary Chronicle (Synodal version)

In this table, the weak pronouns are represented in the columns FamTop, InfoFoc, and Bg, whereas all other assigned discourse functions require strong pronouns. The weak instances amount to only 2 cases (= 5%). One of the latter could still be understood contrastively; the other one might be demonstrative:

(49) Оканьнии же они оттолѣ пришедше, взяша Москву,
accursed-PL MP they from-there come-CVB.PT.PL take-AOR.3PL M.-ACC
Переяславль, [...]
P.-ACC
'The (those?) accursed ones, having arrived from there, took Moscow, Perejaslavl'
[...]' (NovgStar [1238], RRuDi)

A few weak uses of 1/2P pronouns are rather unquestionable, but in general, their information-structural classification is difficult. Consider (50):

(50) [...] а Ярославъ всеѣ злобы лишается, а за то язъ₁
and Ja.-NOM all evil-GEN abandon-3SG-REFL and for that I-NOM
поручаюся; аже будете и крестъ цѣловали, язъ₂ за то прииму
guarantee-1SG that AUX-FUT.2PL also cross kiss-LPT.PL I for that do-1SG
опитемью [...]
penance
'[...] and Jaroslav will abandon all evil and for that I guarantee; that you will kiss the cross, I will do penance for that [...]' (NovgStar [1270], RRuDi)

On the view of information structure applied here, *jazъ₁* should not form an Aboutness Topic of its own (the Aboutness Topic being *za to*). So we would analyse it as "OtherBackgr" and thus a case of *weak* pronoun use. *Jazъ₂* then constitutes a new Aboutness Topic, shifting from *za to*, and thus a *strong* use. It cannot be just "Familiar" because of being referentially easily accesible, because that would hold of just any use of a 1/2P pronoun.

Proponents of Borkovskij's "logical stress" view might certainly find an emphatic overtone of both instances of *jazъ* in (50); but a definition or independent criterion to support this has never been provided. Even worse, it is very plausible that the mere use of a grammatically "superfluous" pronoun kicks off a process of Gricean reasoning: The violation of the maxime of Quantity raises the implicature that the speaker must have associated special importance with *jazъ*. But if this is on the right track, then just *any* such "superfluous" use will create

the impression of emphasis, and the relation between emphasis and the realisation of a weak pronoun becomes fully circular. Much would therefore depend on the pinning down of the concept of emphasis in a way that is independent of this pragmatic effect. In the meantime, we have to admit that there is some grey area even in a null subject language, in which weak 1/2P pronouns can be used without much linguistic motivation.

It must also be noted that 3P does not follow this line at all, in the given text; there are no weak 3P pronons to be found. This is an early indication of a different development according to grammatical person. In any event, it would be very misleading to doubt the pro-drop status of the language of the Novgorod First Chronicle, given the vast quantitative predominance of null subjects over subject pronouns.

Our sample from the oldest available chronicle in the Kiev tradition, the *Povest' vremennyx let* (Laurentius chronicle, early 12th cent., redaction 1377) shows the following distribution of pro-drop across tenses and persons:

	Aorist	Perfect	Past	Present
1/2P	4 (=100%)	2 (=100%)	–	15 (=88.24%)
3P	62 (=82.67%)	1 (=100%)	1 (=100%)	11 (=73.33%)

Table 3.7: Amount of pro-drop in Nestor's chronicle (Laurentius), early 12th c. (603 finite clauses; % of relevant subjects)

Past tenses in this text sample almost exclusively amount to aorists, which show explicit, unambiguous person marking and a high amount of pro drop. Bare LPT forms were virtually non-existent at this point in the development, except for 3P, where the auxiliary was dropped earlier than in the other persons. At the given stage, no clear differences between the persons with respect to verb impoverishment can be deduced due to the very low numbers of perfect and bare LPT examples; but at least the three instances of the bare LPT with pro-drop vs. none with pronominalisation point in the expected direction.

Turning to the information-structural and referential characteristics of overt subject pronouns vs. *pro* in the above sample, it is important to note that the non-null instances of subjects form a mixed bag, containing not only personal pronouns, but also coreferential demonstratives (*cь 'this (one)'*) and even lexical nouns or names representing Familiar Topics, such as the second instance of *Olegъ* in the following example:

(51) И   сѣде   Олегъ   княжа   въ Киевѣ, и   рече   Олегъ: «Се
      and   seat-AOR.3SG   O.-NOM   rule-CVB.PRS   in K.-PRP.SG and say-AOR.3SG   O.-NOM   this
      буди   мати   градомъ   русьскимъ».
      be-IMP.2/3SG   mother   city-DAT.PL   Russian-DAT.PL

'And O. sat ruling in Kiev, and O. said: "This shall be the mother of Russian cities." '

(Lavr.Let., RRuDi)

Overt *onъ* invariably fulfils the information-structural function of a Contrastive or Shifting Topic, and never occurs without the contrasting connectives *i, a* or the adversative particle *že*. A typical usage is to shift back and forth between participants of a conversation, cf.

(52) Она   же рекоста: ``Вѣ вѣвѣ,   како есть человѣкъ
     they-DU.NOM MP said-AOR.3DU know know-know how be-3SG man
     створенъ''. Он же рече: ``Како?'' Она   же рекоста: [...]
     made-PAS.SG he MP said how   they-DU.NOM MP said-AOR.3DU [...]
     'And they said: "Know, how man is made". And he said: "How?" And they said: [...]'

(Lavr.Let., RRuDi)

Interestingly, Topics of the spoken part itself seem to have no influence on the Topic progression; a subsequent subject is still marked as a Shifting Topic if it does not continue the introducing, main clause topic. The same holds for embedded clauses, cf.

(53) Онъ же [...] приде   в манастырь ту,   и умоли   игумена
     he   MP   came-AOR.3SG in monastery this-ACC and asked-AOR.3SG abbot-ACC
     того, дабы   на нь възложилъ   образ мнишьскый. Онъ же
     that-ACC that-SBJ on him laid-LPT.M.SG picture monk-   he   MP
     послушавъ   его,   постриже и, [...]
     listen-CVB.PT shave-AOR.3SG him
     'And he$_i$ [...] came to this monastery and asked that abbot$_j$ to put the monk's picture
     on him$_i$. And he$_j$, listening to him$_i$, cut his$_i$ hair, [...]'   (Lavr.Let., RRuDi)

Investigation of all 3P subject pronouns in the full text yielded the results in 3.8.

	ConTop	ConFoc	FamTop	MinInfoFoc	InfoFoc	OpFoc	Bg	ShiftTop	Sum
3P	4	0	0	2	0	0	0	57	63

Table 3.8: Subject pronouns in Nestor's chronicle (Laurentius redaction)

There are no weak pronouns in 3P, but quite a few in 1/2P, cf.

(54) Доколѣ не насытистеся злобъ   вашихъ? Вы   бо
     still   not be-sated-AOR.2PL wickedness-GEN.PL your-GEN.PL you-2PL because
     уклонистеся от   пути моего,   глаголеть господь [...]
     deviate-AOR.2PL from path my-GEN.SG says   Lord
     'Are you still not sated with your wickedness? Because you have left my way, says
     the Lord [...]'   (Lavr.Let., RRuDi)

(54) might count as "emphatic", but, again, no independent definition of linguistic emphasis has been provided which would fit exactly here. The overt pronoun continues the Topic of the previous clause and thus counts as weak.

The overall situation is thus similar to the one in the earlier chronicle: <mark>Pro-drop abounds and overt 3P subject pronouns are rare; when they occur, they are strong pronouns functioning as contrastive, minimally focused, or as Shifting Topics.</mark> However, we find <mark>several weak uses of 1/2P subject pronouns</mark> which do not lend themselves to any obvious syntactic or information structural explanation.

Epos: Zadonščina The only Old Ru epos, the late 14th c. Zadonščina, mirrors the use of null subjects observed in early chronicles and hagiography. The text is shorter than our usual sample size (only 315 clauses instead of 600).

	Aorist	Perfect	Past	Present
1/2P	3 (=100%)	1 (=50%)	1 (=50%)	22 (=91.67%)
3P	7 (=100%)	–	6 (=100%)	13 (=100%)

Table 3.9: Amount of pro-drop in *Zadonščina*, late 14th c. (n = 315; % of relevant subjects)

Weak 3P pronouns are non-existent, and also 1/2P null subjects prevail over realised pronouns (27 vs. 8 occurrences). Almost all of the latter represent Shifting Topics (3 occurrences) and/or are followed by an apposition (4 occurrences), e.g.

(55) Чему ты, поганый Мамай, посягаешь на Рускую землю?
      why   you   faithless M.   touch-2sg   on Russian soil-acc
      'Why do you, faithless Mamaj, touch on Russian soil?'   (Zad, RRuDi)

The main detectable change does not concern pro drop, but the general distribution of verbal preterite forms: Where AOR was by far predominant earlier, bare and periphrastic LPT forms now exceed it in frequency (53 AOR vs. 80 bare LPT vs. 11 periphrastic LPT overall). This raises the problem of how to properly analyse the LPT.3P forms (given under "Past" in the above table) – as instances of a morphologically zero 3P auxiliary (i.e., *de facto* as "Perfect"), or as dropped auxiliaries (*de facto* bare "Past"). While bare past tense forms were so rare earlier that the assumption of a dropped auxiliary was simply implausible, this is changing now, as <mark>Aorists and Perfect forms are fading away.</mark>

15th century prose

Afanasij Nikitin's *Choždenie za tri morja*   The *Choždenie za tri morja* (1468–1475) is especially interesting in our context, because it contains a variety of preterite forms in parallel, with Aorists already occurring more rarely than the periphrastic Perfect forms or bare LPTs. The text highlights the different developments in 3P vs. 1/2P. As in the *Zadonščina*, 1/2P strongly favour the Perfect (49 instances in our sample) over the bare LPT (14 instances), whereas there are no examples of a periphrastic Perfect (with a realised auxiliary) in 3P. Pro-drop shows a clear connection to auxiliary realization in 1/2P, with all Perfect examples, but only a relatively small amount of bare LPTs lacking an overt subject.

	Aorist	Perfect	Past	Present
1/2P	14 (=100%)	42 (=100%)	2 (=100%)	10 (=100%)
3P	–	–	26 (=92.86%)	57 (=90.48%)

Table 3.10: Amount of pro-drop in *Choždenie za tri morja*, 15th c. (n = 611; % of relevant subjects)

Note that the 100% pro-drop throughout 1/2P does not mean that there are no subject pronouns in the text, but only that they function as strong pronouns throughout, and are thus not comparable to *pro*. The amount of pro-drop in 3P is high, but slightly reduced vis à vis the older texts.

We cannot tell apart the periphrastic Perfect (assuming a zero morpheme as auxiliary) from the bare LPT in 3P in principle, but the fact that pro-drop still abounds in PRS.3P and in 1/2P tells us that it had to be fully grammatical at least in tenses with explicit agreement morphology. Since neither Perfect nor Aorist were obviously needed in 3P (there are only 2 AOR.3P in our sample, none of them with pro-drop), the most plausible explanation is that the bare LPT in 3P could still be understood as a morphologically explicit verb form, involving a meaningful zero auxiliary – just as in the chronicles or early charters.

Consequently, pro-drop with a bare LPT must have had a different status in 1/2P vs. 3P. Only in the former case may we plausibly speak of verb impoverishment, while in the latter, the default marker for person had early become a null morpheme, which was, however, still fully informative for grammatical person in mid 15th c. (56) illustrates the typical patterns. In (56-b), pro-drop in bare LPT.3P even relates to an object antecedent across several clause boundaries, with different intervening topics, demonstrating once more that pro-drop was clearly less restricted overall than in modern Ru (cf. section 3.4.2).

(56) a. А   писание   то   своею   рукою   написал
          and writing-ACC this-ACC POSS.REFL-INS.SG hand-INS.SG wrote-PT.M.SG
          'And this writing, I wrote with my own hand.'

      b. [И яз грешный привезлъ жеребца*ᵢ* в Индийскую землю, и дошелъ есми

          до Чюнеря богъ далъ поздорову все, – 'And I, sinner, brought a stallion to

          India, and we made it to Chjuner', God gave welfare,']

          a   *proᵢ* стал   ми   во   сто рублев
          and   cost-LPT.M.SG   me-DAT   in   100 roubles-GEN.PL
          'and it (= the stallion) cost me about 100 roubles.'   (Nikitin, RRuDi)

An analysis of subject pronouns in the full text (rather than only the 600 clauses sample)
points to the same generalisations: I could identify 15 occurrences of strong 3P pronouns vs.
3 weak ones – yet a small number, but not a complete lack, as was the case in *Zadonščina*
and most of the chronicles.

The Tale of Dracula (*Povest' o Drakule voevode***)**   The *Tale of Dracula*, which also stems
from the second half of the 15th century, presents a more old-fashioned pattern of verbal
forms than the *Choždenie za tri morja*, with many AOR.3P and even some 3P auxiliaries (pe-
riphrastic Perfect forms). In detail, we find 172 Aorists, 4 Perfect, and only 15 bare LPT forms
in 3P overall. Other archaisms include *dativi absoluti* like

(57) Умершу   же тому   воеводе,   и   краль   пусти   к   нему   в
      die-CVB.PT.DAT MP this-DAT.SG duke-DAT.SG and king-NOM.SG allow   to him-DAT in
      темницю   …
      dungeon-ACC
      'But when this ruler had died, the king let (someone) to him into the dungeon …'
                                                                      (Drak, RRuDi)

Our sample includes the full text, which amounts to some 360 clauses. Table 3.11 shows
the distribution of pro-drop in present and preterite:

	Aorist	Perfect	Past	Present
1/2P	14 (=100%)	8 (=100%)	– (1 pron.)	21 (=95.45%)
3P	95 (=98.96%)	2 (=100%)	6 (=85.71%)	18 (=90%)

Table 3.11: Amount of pro-drop in *Povesť o Drakule voevode*, 15th c. (n = 360; % of relevant
       subjects)

In two respects, this overview ties in with the previous sample: First, there is again a slight
reduction of pro-drop vs. subject pronominalisation in Present tense, notably in 1/2P, but

pro-drop is obviously fully grammatical and widely used. Second, the connection between the realisation of an auxiliary and pro-drop is still obvious in 1/2P preterite forms. Given the latter observation, it seems again plausible that the bare LPT involves a morphologically null 3P auxiliary, in which case the high amount of pro-drop with LPT.3P would not come as a surprise. The clear difference between the *Choždenie za tri morja* and the *Povest' o Drakule voevode* with respect to the occurrence of 3P Aorist and Perfect forms shows that, in texts of the same period and similar genre, pro-drop was independent of the choice of a specific preterite tense in 3P, whereas it was strongly correlated with this choice in 1/2P. Again, this is plausible if "Past" tense contains a 3P null auxiliary, in which case all preterite forms are equally explicit with respect to agreement morphology.

Regarding the properties of subject pronouns, while 3P pronouns are invariably strong (30 Contrastive or Shifting Topics), a small amount of weak 1/2P pronouns can be found (3 vs. 15 strong uses), cf.

(58) Право рекл   еси.   ты   еси   велика   государя
      true   say-LPT.M.SG AUX-PRS.2SG you-NOM be-PRS.2SG great-GEN.SG lord-GEN.SG
      посол   кралевьскы [...]
      legate-NOM king's-NOM.SG
      '(You) said the truth: You are a royal legate of a great lord [...]'   (Drak, RRuDi)

This distribution is in line with the more archaic nature of this text than the *Choždenie za tri morja*, which does have single weak 3P pronouns.

16th/17th century

Domostroj   The *Domostroj* consists of a mixture of renditions of several older sources concerning a proper Christian everyday life, with more practical advice, from cooking and brewing recipes to the right clothing. While the former parts are strongly influenced by Church Slavonic, this is less obvious for the latter. Although we took care to include an equal amount of both subparts in our sample, the result is very homogeneous and close to the picture provided by e.g. the chronicles or the *Tale of Dracula*: Pro-drop is the rule across all tenses and persons, with next to no exceptions.

	Aorist	Perfect	Past	Present
1/2P	17 (=100%)	5 (=100%)	2 (=100%)	25 (=100%)
3P	1 (=100%)	25 (=100%)	11 (=100%)	26 (=96.3%)

Table 3.12: Amount of pro-drop in *Domostroj*, mid 16th c. (n = 639; % of relevant subjects)

Note that the redactions of *Domostroj* vary somewhat in the distribution of preterite verbal forms, in respects relevant for us:

(59) a. тѣмъ   всегда оутешⷶ̆   [утешал]   себѣ   ... (RRuDi)
        this-INS.SG always consolate-AOR.1SG consolate-LPT.M.SG self-DAT

   b. Тѣмъ   всегда утешал   себѣ   ... (PD)
        this-INS.SG always consolate-LPT.M.SG self-DAT
        'I always consolated myself by the fact that ...'   (Domostroj)

The first variant in (59-a) would count as morphologically explicit agreement, the latter and the one in (59-b) would not.

An in-depth analysis of all subject pronouns in the full text reveals only one weak 3P and a handful of 1/2P subject pronouns, as in

(60) не вѣдаю   азъ ничего того   и не слыхалъ   и не знаю
      not know-1SG I nothing that-GEN.SG and not hear-LPT.M.SG and not know-1SG
      'I do not know anything about that, nor did I hear or know anything.' (Domostroj, RRuDi)

Nevertheless, the quantitative picture is extremely clear: There is no indication of any reduction or loss of pro-drop.

The Life of Avvakum (*Žitie protopopa Avvakuma*)   The *Žitie protopopa Avvakuma* (2nd half 17th c.) is an autobiographical narrative text, again with a characteristic mixture of registers, reaching from jargon and *prostorečie* to elevated literary (religious) passages. The periphrastic perfect has virtually died out, possibly with the exception of single examples in the stylistically elevated parts of the text, and has been fully replaced by the bare LPT. Pro-drop is still frequent, but weak subject pronouns are clearly gaining ground.

	Aorist	Perfect	Past	Present
1/2P	5 (=100%)	1 (=100%)	3 (=60%)	35 (=79.55%)
3P	6 (=100%)	–	47 (=90.38%)	35 (=72.92%)

Table 3.13: Amount of pro drop in Avvakum's *Žitie*, 2nd half 17th c. (n = 631; % of relevant subjects)

The different diachronic development of pro-drop with respect to 1/2P vs. 3P continues in Avvakum's *Žitie*: In Present tense, it becomes rarer for all persons; in preterite tenses, there is a slighter overall decrease. But mainly, pro-drop occurs with both AOR and bare LPT in 3P, but much more rarely with LPT.1/2P. This points to the generalisation that there is still a null

auxiliary active in LPT.3P, and there is still a connection between auxiliary drop and pro-drop in 1/2P.

It must be noted that as before, all instances of pronouns as Shifting Topics have been removed from consideration; otherwise, the amount of subject pronominalisation – especially in 1/2P – would be much higher. This also excludes rather frequent borderline cases like (61) or (62):

(61) [Topic ≠ 1P …] И   я   говорю: [...] / И   я   закричал: [...]
                      and I-NOM say-1SG   and I-NOM scream-LPT.M.SG
      '[Topic ≠ 1P …] And I said: [...] / And I screamed: [...]'   (Avv., RRuDi)

(62) Так он   велел   перестать. И я   промолыл ему: «за что
      so he-NOM order-LPT.M.SG stop   and I-NOM tell-LPT.M.SG he-DAT for what
      ты   меня   бьешь? ведаешь ли?» И он   паки велел   бить
      you-NOM me-ACC beat-2SG know-2SG QU   and he-NOM then order-LPT.M.SG beat
      по бокам,   и   отпустили. Я   задрожал,   да   и   упал.
      on side-DAT.PL and leave-LPT.PL I-NOM tremble-LPT.M.SG and even fall-LPT.M.SG
      'So he gave the order to stop. And I asked him: "Why do you beat me? Do (you) know?" And he again gave the order to beat by the sides, and (they) left me. (I) trembled and even fell.'   (Avv., RRuDi)

In these examples, there is no explicit indication of a topic shift (like the particle *že* or adversative *a, no*). Nevertheless, all occurrences of subject pronouns with the exception of *ty* – which is not an Aboutness Topic in the question context – have been classified as Aboutness-Shift Topics. This seems justified because the use of a subject pronoun in such an information-structural constellation is not indicative of a non-null subject language at all: Italian, Czech, or Polish, among many other canonical null subject languages, use a (strong) pronoun to mark a topic shift. The fact that *Avvakum* contains many more subject pronouns than, e.g., the *Domostroj* (namely, 501 out of 25110 tokens vs. 68 out of 27356 tokens), does not tell us much by itself. However, there is indeed a certain – possibly individual, stylistic – tendency for Avvakum to pronominalise the 1P subject, even where it is a background element, cf.

(63) протопоп, ведаю   -де я   твое   чистое   и   непорочное
      p.-NOM   know-1SG MP I-NOM your-ACC.SG clean-ACC.SG and pure-ACC.SG
      житие
      life-ACC.SG
      'Protopope, I know your clean and pure conduct of life.'   (Avv., RRuDi)

Note that formulae like the frequent, religiously motivated *az" grešnyj 'lit. I sinner / The*

sinner I am' were also excluded from the above calculations, because they cannot alternate with *pro* anyway.

Although the change is not yet fundamental, subject pronouns do increasingly become used as weak pronouns instead of former *pro*, even in tenses with explicit marking of person agreement (present tense). It is important to note that this also affects 3ᴘ pronouns now, as the distribution of information-structural functions over 3ᴘ pronouns in the full text shows (see table 3.14).

	ConTop	ConFoc	FamTop	MinInfoFoc	InfoFoc	OpFoc	Bg	ShiftTop	Sum
3ᴘ	2	0	12	2	0	1	1	102	120

Table 3.14: Subject pronouns in Avvakum's *Žitie*

Weak uses (*FamTop, NonMinInfoFoc* and *OtherBackgr*) make up 13 instances of the pronoun *on* (10.83%).

Early 18th century

The Tale of Frol Skobeev (*Povest' o Frole Skobeeve***)**   In 18th c. prose texts like the popular *Tale of Frol Skobeev*, recorded around 1720, the system of preterite tenses has been completely reduced to the bare ʟᴘᴛs. At this point in the development, a considerable decrease of pro-drop has taken place across persons and tenses (table 3.15).

	Aorist	Perfect	Past	Present
1/2ᴘ	–	1 (=100%)	2 (=18.18%)	25 (=71.43%)
3ᴘ	–	–	68 (=54.84%)	6 (=24%)

Table 3.15: Amount of pro drop in *Frol Skobeev*, 1720 (n = 642; % of relevant subjects)

It should be noted that not only weak pronouns, but also full noun phrases which repeatedly refer to the same protagonist (Familiar Topic) and thus alternate with *pro*, are typical for this text. Nevertheless, pro-drop is still preferred over subject pronominalisation in some tenses/persons, e.g. in present tense 1/2ᴘ.

Interestingly, pro-drop in 1/2ᴘ with bare ʟᴘᴛs – the constellation which was virtually excluded in earlier texts – only occurs in the low variety, colloquial utterances of the coachman in (64). This may already point towards a stylistic differentiation which is typical for the situation in modern Ru.

(64) толко *pro*$_i$ помню,   какъ *pro*$_i$ приехалъ   к нему на квартиру,
     only   remember-PRS.1SG how   came-LPT.M.SG to he-DAT in flat-ACC.SG
     а   куды онъ Скобеев$_j$ ездилъ   и   что *pro*$_j$ делалъ, *pro*$_i$ не
     but where he S.-NOM   drive-LPT.M.SG and what   do-LPT.M.SG   not
     ведаю
     know-PRS.1SG
     '(I) only remember how (I) drove to him to his flat, but where he, Skobeev, went and
     what (he) did, (I) don't know.'   (Frol, RRuDi)

3P pro-drop alternates with weak pronouns in positions in which the latter used to be avoided earlier, such as the subject of a subjunctive complement clause:

(65) Потомъ Фролъ Скобеев сталъ   просить того столника, чтобъ
     then   F.   S.-NOM began-LPT.M.SG ask-INF this sewer-ACC that-COND
     онъ   пожаловалъ ему корету   и   з   возниками
     he-NOM lend-LPT.M.SG he-DAT coach-ACC also with coachman-INS
     'Then Frol Skobeev began to ask this sewer to lend him his coach, together with a
     coachman.'   (Frol, RRuDi)

None of our earlier text samples contained an overt pronoun in this environment; the use of a null subject instead is illustrated in (66), from the *Choždenie za tri morja*:

(66) и   били   есмя ему$_i$ челом,   чтобы *pro*$_i$ нас   пожаловалъ,
     and beat-LPT.3PL aux-1PL he-DAT forehead-INS that-COND   we-ACC bless-LPT.M.SG
     чем дойти до Руси.
     than get-INF to R.
     'and we bowed down before him, for (him) to bless us before going to Russia'
                                                                   (Nikitin, RRuDi)

The conclusion is thus twofold. First, pro-drop is losing ground, albeit still being well-attested. Second, the conditions on its application have changed: It is now available with bare LPT forms also in 1/2P, depending on stylistic rather than grammatical constraints; and, more importantly, weak pronouns become an alternative in an increasing number of environments. Accordingly, they amount to almost one third (10 out of 31) of the realised 3P subject pronouns, cf. table 3.16.

	ConTop	ConFoc	FamTop	MinInfoFoc	InfoFoc	OpFoc	Bg	ShiftTop	Sum
3P	0	0	3	0	0	0	7	21	31

Table 3.16: Subject pronouns in *Frol Skobeev*

3.5.1.3.3 Diachronic analysis The development of the relative amounts of pro-drop, as discussed above, has been plotted for illustration in Figures 3.1-3.4 below. The 3P Perfect was left out for reasons of sparseness of evidence.
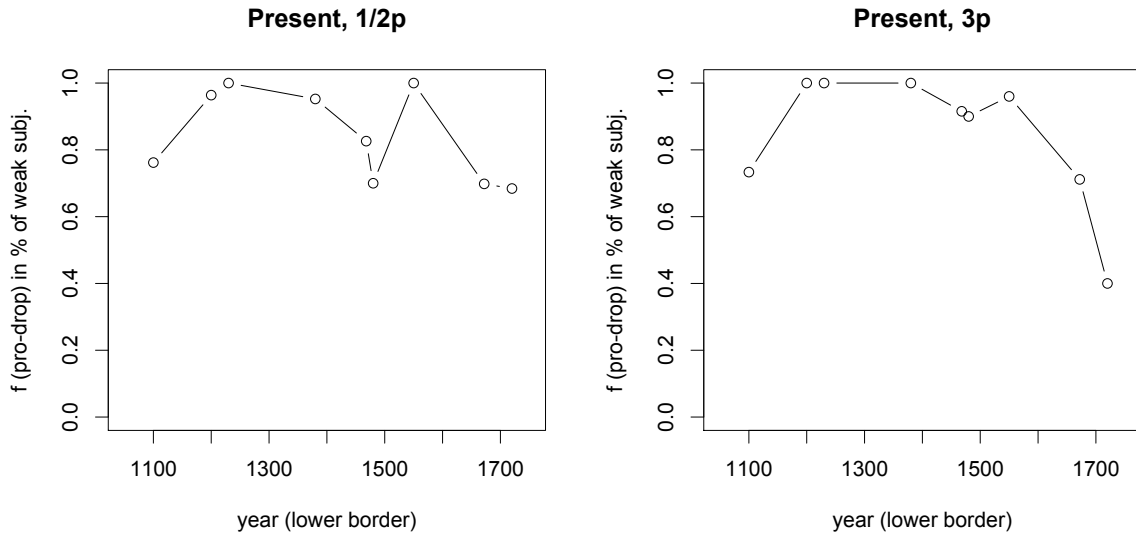

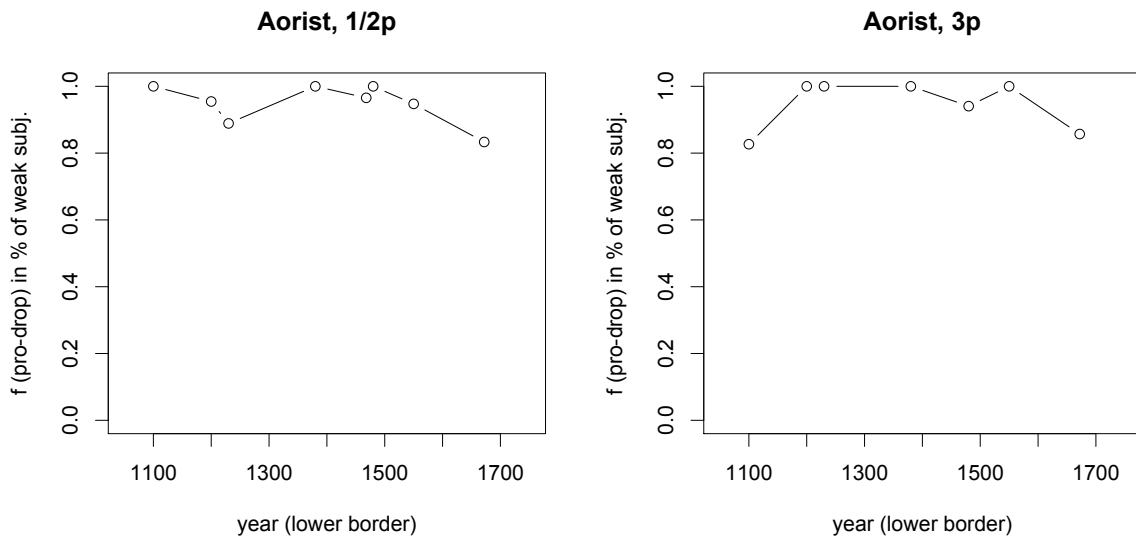Figure 3.1: Overview: Amount of pro-drop in Present tense, samples 12th – 18th c.

Figure 3.2: Overview: Amount of pro-drop in Aorist, samples 12th – 18th c.

Past, 1/2p   **Past, 3p**
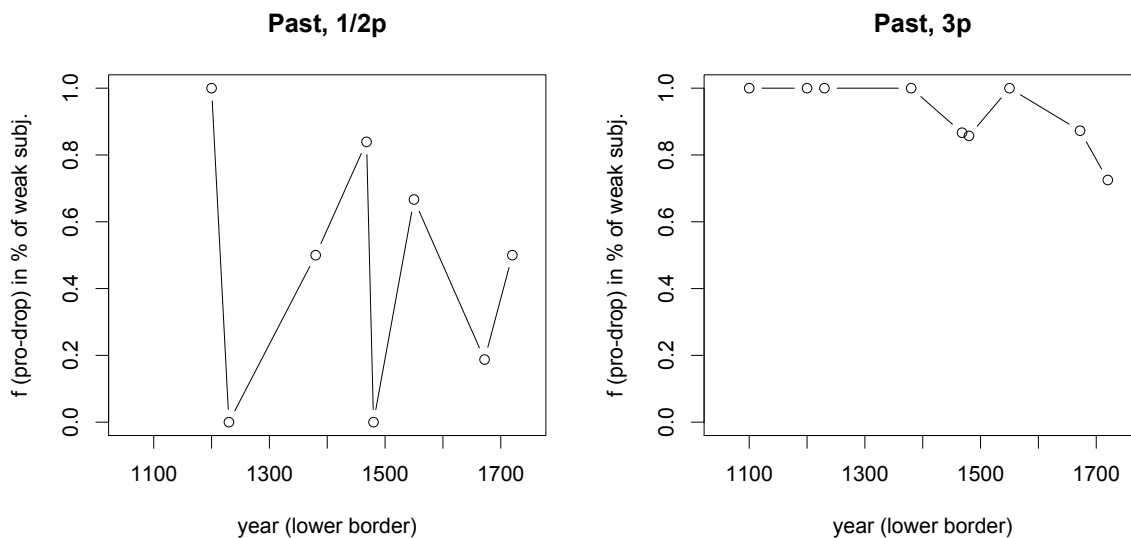
Figure 3.3: Overview: Amount of pro-drop in bare Past tense, samples 12th – 18th c.

Perfect, 1/2p
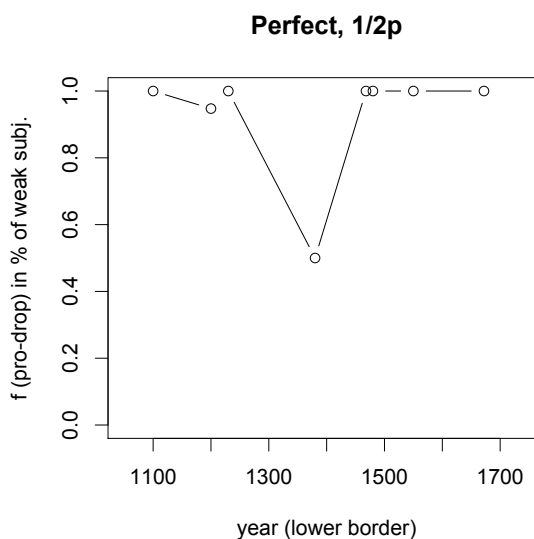
Figure 3.4: Overview: Amount of pro-drop in Perfect, samples 12th – 18th c.

It is certainly misleading that this representation does not also indicate the *absolute* amount of occurrences of a given tense and person. They look as follows (fig. 3.5 for the whole sample corpus, fig. 3.6 only for the cases with weak subjects):

124

Distribution of tense forms, 1/2p

Distribution of tense forms, 3p
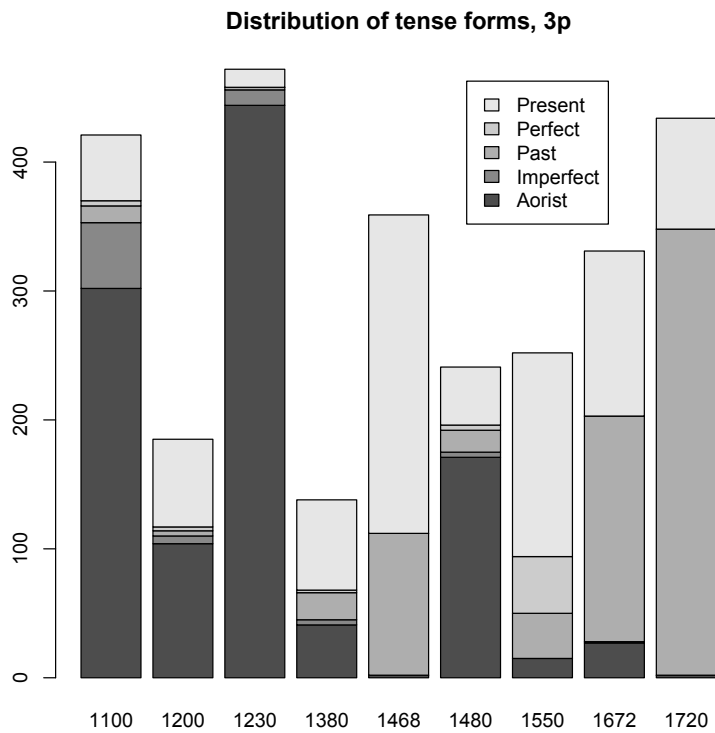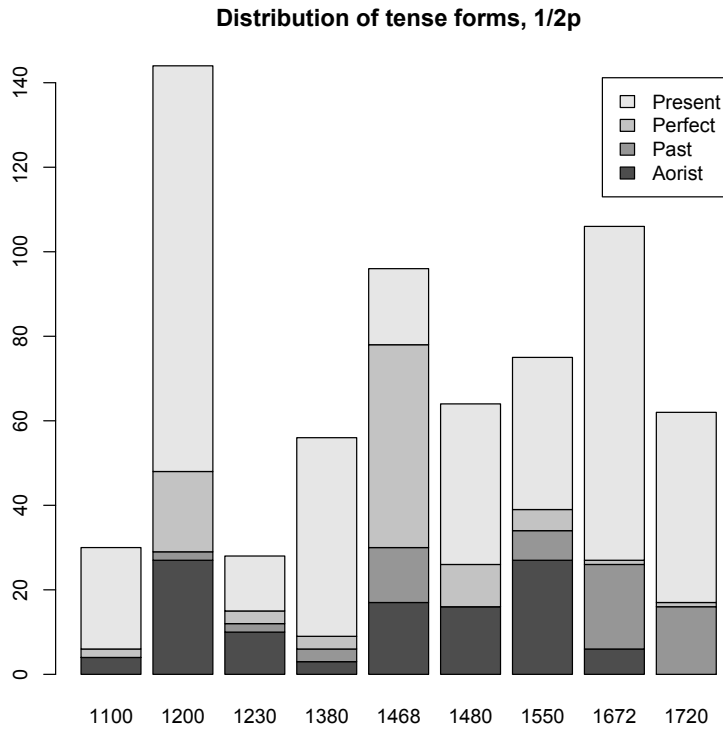
Figure 3.5: Distribution of tense forms, all samples

Distribution of tense forms, 1/2p (only weak subjects)

Distribution of tense forms, 3p (only weak subjects)

Figure 3.6: Distribution of tense forms, all samples

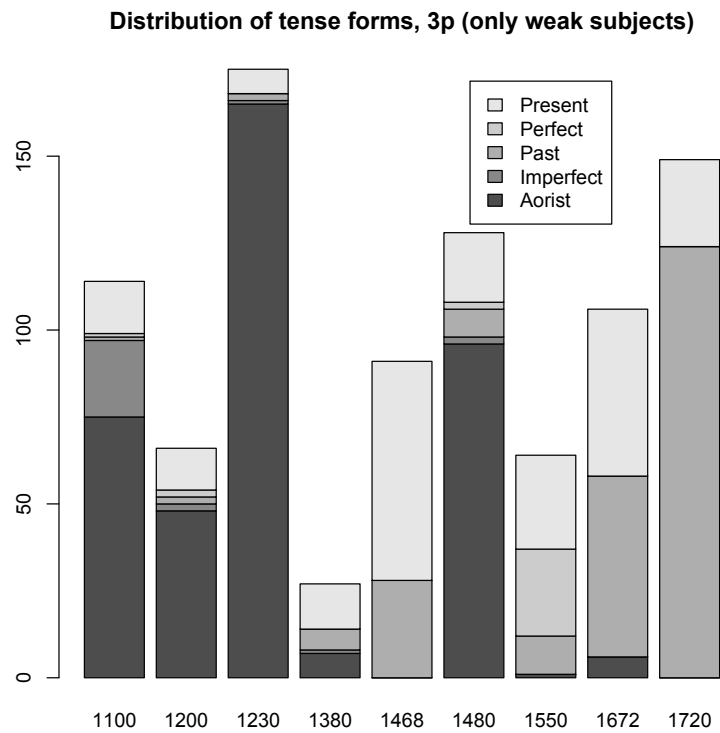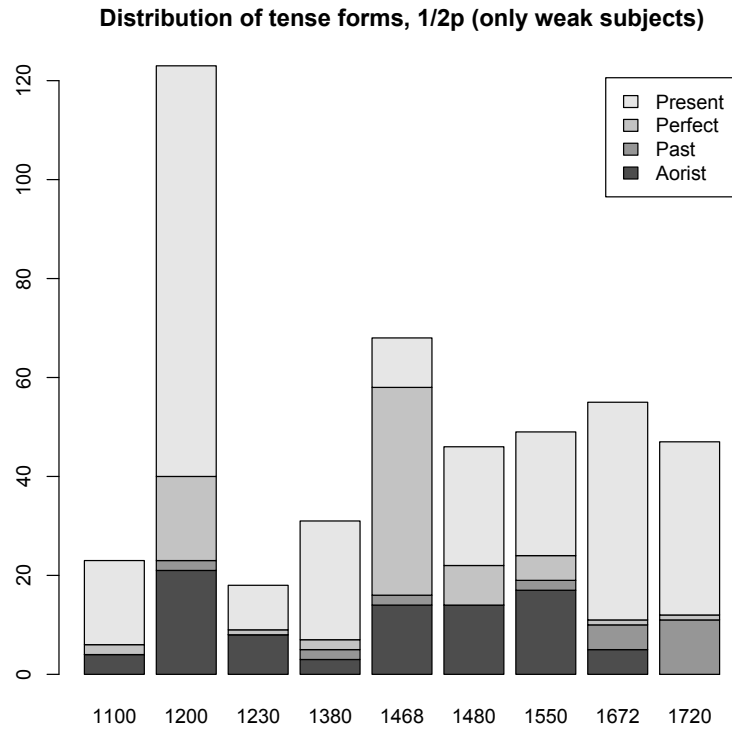It is obvious that – both for the full samples and only for the weak subject cases – Aorist and Perfect diminish overt time, albeit unsteadily. In 3P, Perfect tense (i.e., with an explicit auxiliary) is virtually non-existent. Aorist fades away in the 15th-16th c., with variation between individual texts, and the bare LPT tense form fully replaces it. In 1/2P, the mid 16th c. sample still contains a fair amount of Aorists, but the tendency is similar. The loss of the Perfect auxiliaries, according to the given corpus, occurred during the 16th century.

As we had already seen in the frequency tables above, the very small amount of examples for certain tenses makes it difficult to assess the amount of pro-drop there with any reliability. This holds especially for the 1/2 bare Past tense in earlier samples and the Aorist in later samples. Therefore, the huge variation illustrated in the left part of fig. 3.3 is to be taken with a grain of salt, as is the seemingly high amount of pro-drop in Perfect and Aorist later on. What *is* recognisable, however, is an overall decline of pro-drop from 15th to 17th century. The most important outlier – characterised by an unusually high amount of pro-drop still in mid 16th century – is the *Domostroj*, which does not quite come as a surprise, considering the large amount of older, Church Slavonic texts which entered its composition. This once more points to the importance of the genre factor.

At the level of single examples, subject pronominalisation and the realisation of a Perfect auxiliary are nearly in complementary distribution in 1/2P (cf. Zaliznjak 2004). However, the actual prediction of the verb impoverishment hypothesis is a lot stronger: The change towards the impoverished, bare Past tense form is supposed to have influenced the option of pro-drop for the language as a whole, i.e., across all tenses. If this is true, we should find evidence for a parallel development between the former and the latter (a "constant rate effect", Kroch 2000) even when looking not only at pro-drop in Past tense, but just as well in Present tense, or generally, in all tenses. The diagrams below illustrate various explications of this claim: Fig. 3.7 (left) shows Pro-drop in Present tense (solid line) along with the amount of impoverishment in all preterite tenses (dashed line); fig. 3.7 (right) compares pro-drop across all tenses to the latter.

This is probably the most plausible way of viewing the effect of impoverishment: Among the available preterite forms, the bare, impoverished one is chosen. The extent to which this happens supposedly influences the availability of pro-drop in the grammar as a whole. It is of secondary importance, what amount of *all* verb forms gets inflectionally impoverished, because this number depends on the amount of preterite vs. Present tenses, a specific textual feature outside grammar proper. It is interesting to see whether the course of development was similar for all persons or person-specific. Fig. 3.8 compares 1/2P vs. 3P pro-drop (solid line) to preterite impoverishment (dashed line).
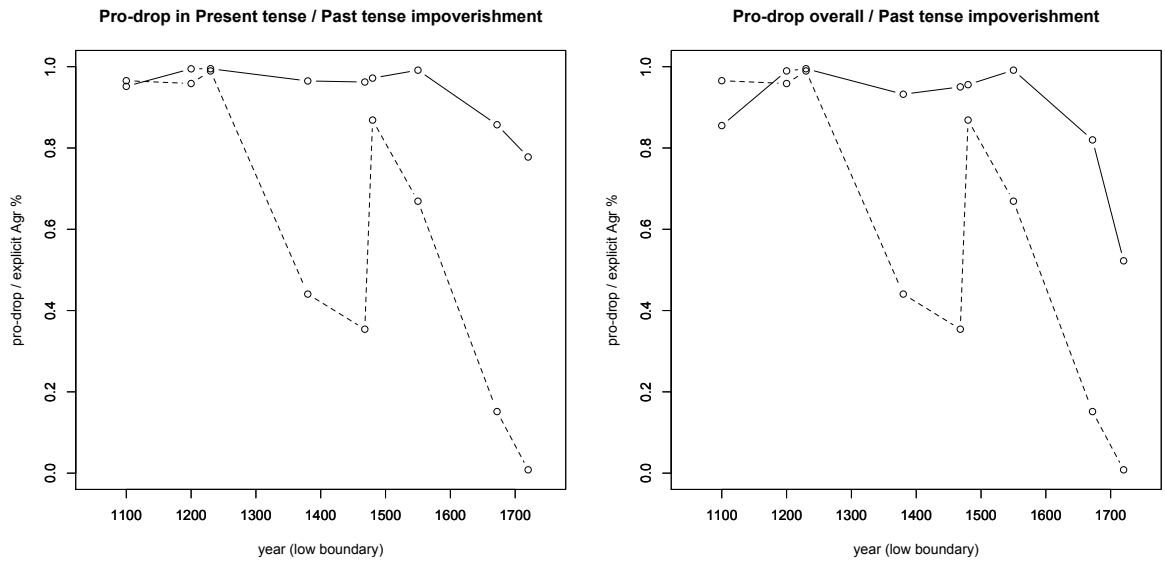
Figure 3.7: Pro-drop in Present tense / in all tenses vs. preterite impoverishment
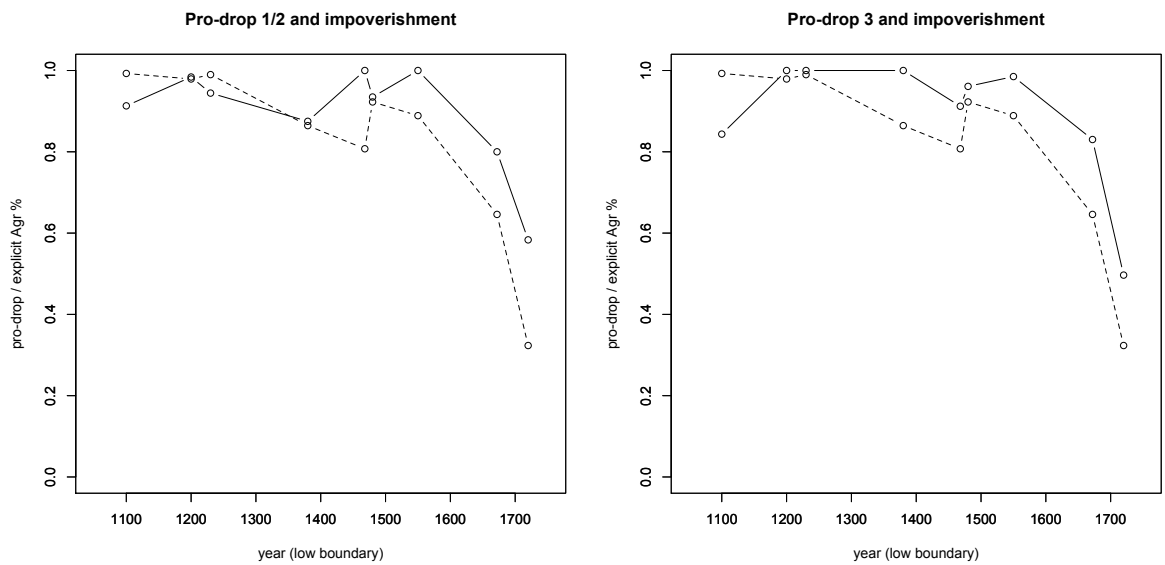

Figure 3.8: Pro-drop in 1/2P / in 3P vs. preterite impoverishment

Although both correspondences are imperfect, it seems that 1/2P essentially deviates in only one data point (mid 15th c.), whereas 3P pro-drop follows the "impoverishment line" less clearly, especially in the middle period (14th-15th c.). This ties in with our assumption that 3P pro-drop in Past tense was indeed supported by a null auxiliary and thus more independent from verbal agreement, at least until the bare LPT established oneself completely.

Kroch (1989) – building on earlier sociolinguistic work by William Labov and others – develops a statistically sound way of dealing with diachronic data of the kind we are facing here. On several examples, including chiefly the rise of *do*-support in the history of English, he demonstrates the method of *logistic regression* for modelling the influences of various independent variables (in our case, mainly *time*) on the choice made by a binary dependent variable (in our case, pro-drop or not, and auxiliary realisation or not, respectively). The major linguistic *tenet* of Kroch's approach is the discovery of so-called *Constant Rate Effects*: If a linguistically motivated theory holds that two phenomena are related – and should be explained by one common construct – then it is natural to assume that these phenomena change diachronically at the same "speed" or rate. Thus, e.g., if the various syntactic configurations in which *do*-support developed, are all related theoretically by one common denominator, then it would be expected that the rate of *do*-support in all of them changed over time *in parallel*.

It is obvious that the verb impoverishment hypothesis makes exactly this kind of claim for Ru: Changes in the past tense paradigm should have occurred in parallel with and at the same rate as the reduction of pro-drop in all tenses. It is important to note that this is a qualitatively different claim than the complementary distribution of 1/2P pronouns and auxiliaries at the level of single examples, because it affects the whole grammatical system across tenses.

Baayen (2008, 214ff) demonstrates in detail how a logistic regression analysis of the kind Kroch (1989) proposed may be done in the statistics language R. Applying the methods described there on our data set, we can now deal with the massive variation in absolute numbers deplored above without problems. As Baayen (2008, 196) writes: "[...] proportions don't provide information about how many observations went into the calculation of the proportion [...] What needs to be done is to weight the proportions for the numbe rof contributing observations." This is where logistic regression steps in by applying a so-called logit transformation on the raw frequency values.

As a first illustration of the results, consider the regression on "auxiliary drop" in 1/2P in our samples, illustrated in the left portion of diagram 3.9.

While the 3P regression curve does not make much sense at all (which is not suprising),
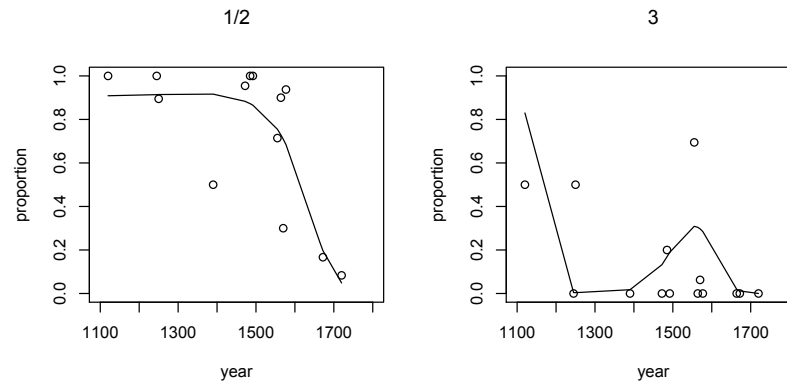
Figure 3.9: Auxiliary drop in past tense(s), logistic regression

the 1/2ᴘ diagram shows the well-known S-shaped timecourse typical of parametric changes which permeate through a corpus and/or population (cf. Kroch 1989 and Roberts 2007a for discussion). According to this regression line, the auxiliary was dropped rather consistently between 1500 and 1700.

In the next step, we would expect that this regression line run in parallel with the decrease in pro-drop across tenses. Let us first consider pro-drop separately for tenses, as we did above. Diagram 3.10 illustrates the regression on pro-drop, conflating grammatical persons into one.

As is apparent, bare Past and Present tense show a clear and consistent decrease of pro-drop; no such development is visible in Aorist, which, however, is less important, because of the general lack of data for the relevant time period. Diagrams 3.11 vs. 3.12 indicate that it was mainly in 3ᴘ that a smooth and consistent decline of pro-drop took place (disregarding the Aorist and Perfect here for lack of evidence in the relevant periods). 1/2ᴘ shows some such development in present tense, but obviously, the development in Past and Perfect was completely unorderly, and – as we know – guided by a different factor, the realisation of the auxiliary.
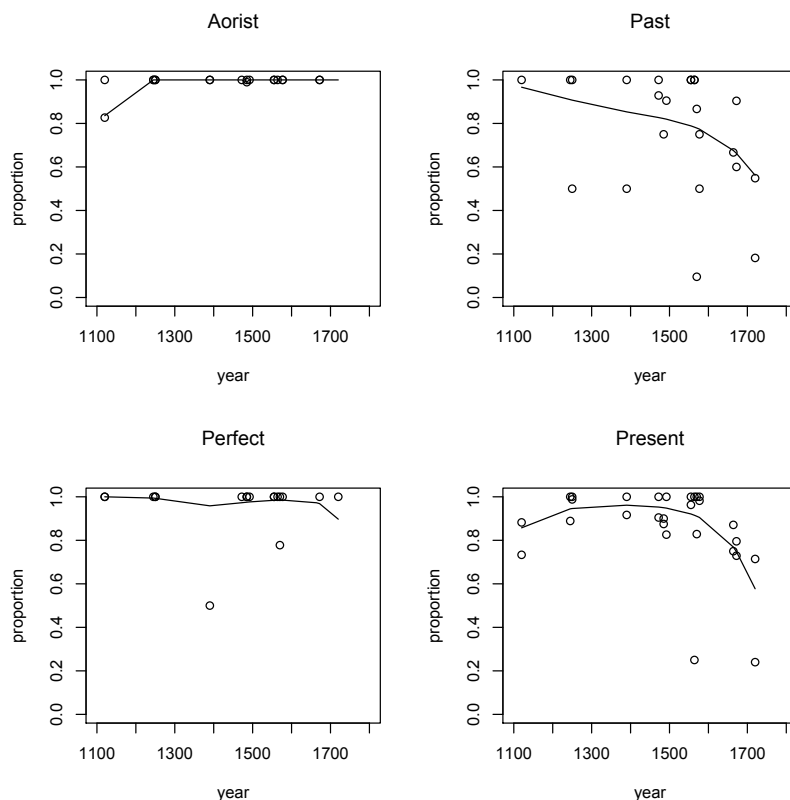
Figure 3.10: Pro-drop per tense, logistic regression

Finally, diagram 3.13 shows an overlay of the regression lines for 1/2P auxiliary loss and for the decrease of pro-drop across persons and tenses. This amounts to the predictions of the verb impoverishment hypothesis in its strict sense.

The regression lines for auxiliary loss (dotted line) and for pro-drop (solid line) quite clearly run in parallel, at least as far as our data goes – more samples from the 18th c. would clearly be advisable to strengthen the result. However, the indication of a Constant Rate Effect is obvious. I conclude that the verb impoverishment hypothesis is, overall and in its strong form, on the right track and supported by our Ru diachronic samples. What was indicated, with a fair amount of "background noise", in the proportion plots above, can be shown, in a statistically sound approach, via logistic regression.

If the verb impoverishment analysis is on the right track, the conflation of all person markers in the past tense paradigm is responsible for the change from a canonical pro-drop language to a partial pro-drop language. An immediate problem for such an explanation – yet overlooked in the relevant literature, be it traditional or recent formal – is that there were systematic person syncretisms already in OCS, and later in ORu verbal inflection: Across all
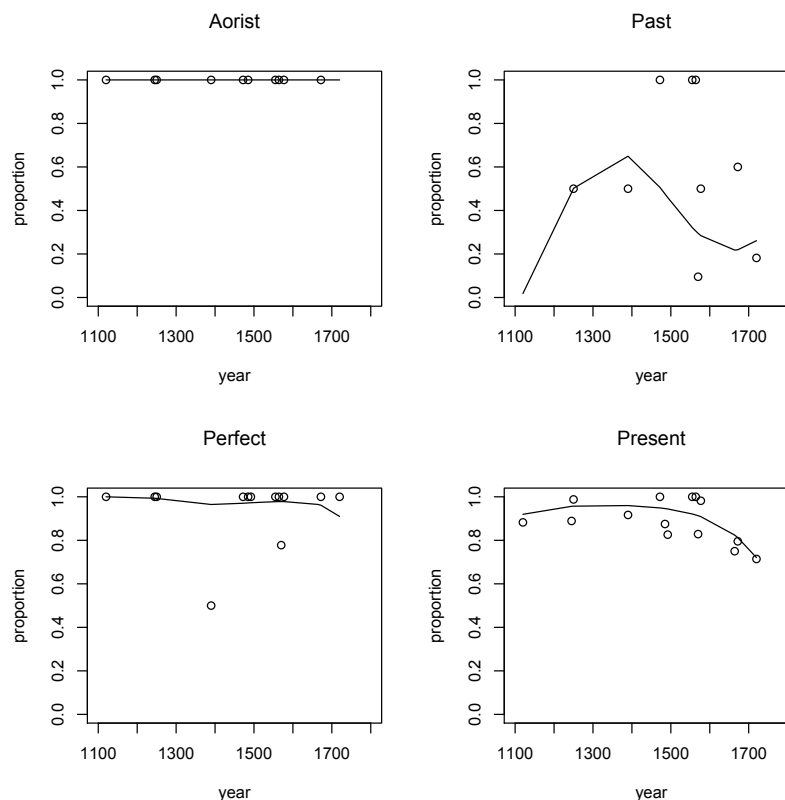
Figure 3.11: Pro-drop in 1/2P per tense, logistic regression

paradigms, the (sigmatic and asigmatic) aorist and imperfect 2sg and 3sg had the same ending (*-Ø*); furthermore, 2PL and 3DU looked the same in all tenses (*-V-te*). Note that aorist was by far the predominant preterite tense in OCS and early ORu "high variety" texts (though virtually absent from the birchbark charters), so the person syncretism in aorist might even have been more relevant in terms of frequency than the change from the rather rare perfect to the bare past tense. In the sense of Jaeggli and Safir (1989), it should have constituted evidence for a non-uniform agreement system, and thus, a non-pro-drop language. But in that case, little would have changed from early Old Ru up until today.

Under Müller's (2006, 2007) substantially improved version of the verb impoverishment hypothesis, however, verbal syncretisms only have an impact on pro-drop properties if they are derived via a *morphological rule* of person impoverishment; so one might argue that this is not the case in OCS and early ORu, leaving their pro-drop status intact. Exactly this type of potential counterevidence is treated in considerable detail by Müller (2007). On the one hand, he shows that German and Icelandic both have systematic 2/3P syncretisms (in present tense), which are based on natural classes of "non-first-person" ([−1]) markers, and concur
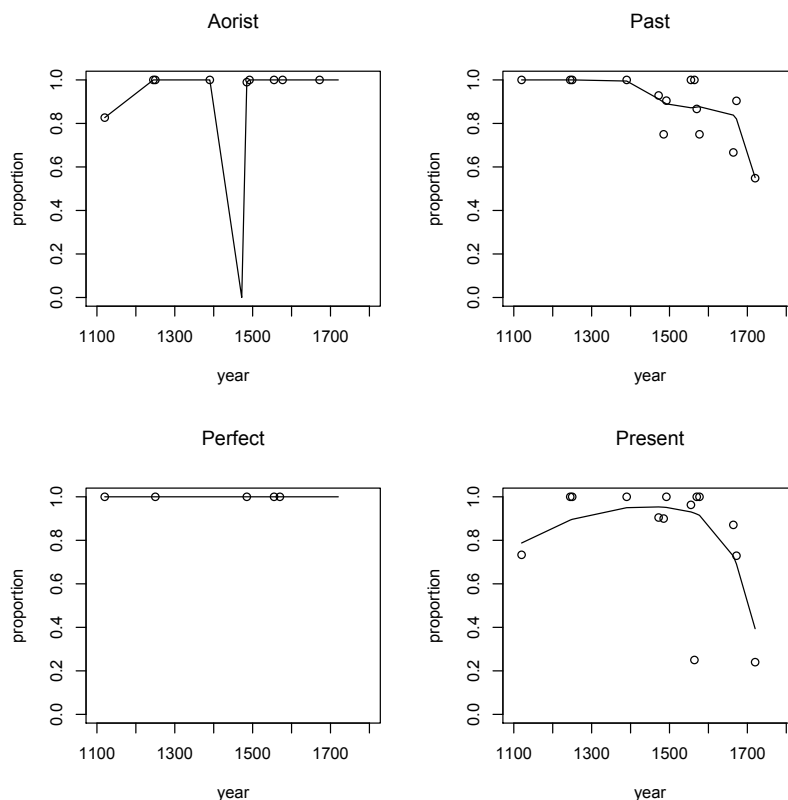
Figure 3.12: Pro-drop in 3P per tense, logistic regression

with a non-pro-drop status. On the other, the two languages Wambon and Kenuzi-Dongola, both discussed in Cysouw (2001) , also have 2/3P syncretisms, but only Wambon is a pro-drop language. Müller argues that Wambon has no (other) system-wide instances of syncretism in its verbal paradigms, and that null markers (-Ø, also the relevant morpheme in Wambon) tend to correlate with radical underspecification, rather than with rules of impoverishment, cross-linguistically. So it would be more adequate to derive the Wambon pattern by assuming a person-specific [+1, −2] morpheme *-ep-* and a morpheme *-Ø*, underspecified completely for person ([ ]). Since only the most specific morpheme fitting the given environment is selected, *-Ø* will be overridden by *-ep-* in 1P. By contrast, Kenuzi-Dongola shows 2/3P syncretisms in various tenses, involving several different, non-null markers. In this case, a rule removing person feature specifications and leaving only [−1], i.e., an impoverishment rule, is called for.

It seems justified to assume that OCS and early ORu 2/3P.AOR/IMPF syncretisms are rather of the first (Wambon) type than of the latter (Kenuzi-Dongola). First, they only involve null endings. Second, 2/3P are systematically distinguished in present tense, so the syncretisms are somewhat exceptional rather than system-wide. Note that the 2PL/3DU syncretism in-

AUX drop (1/2p) and pro-drop
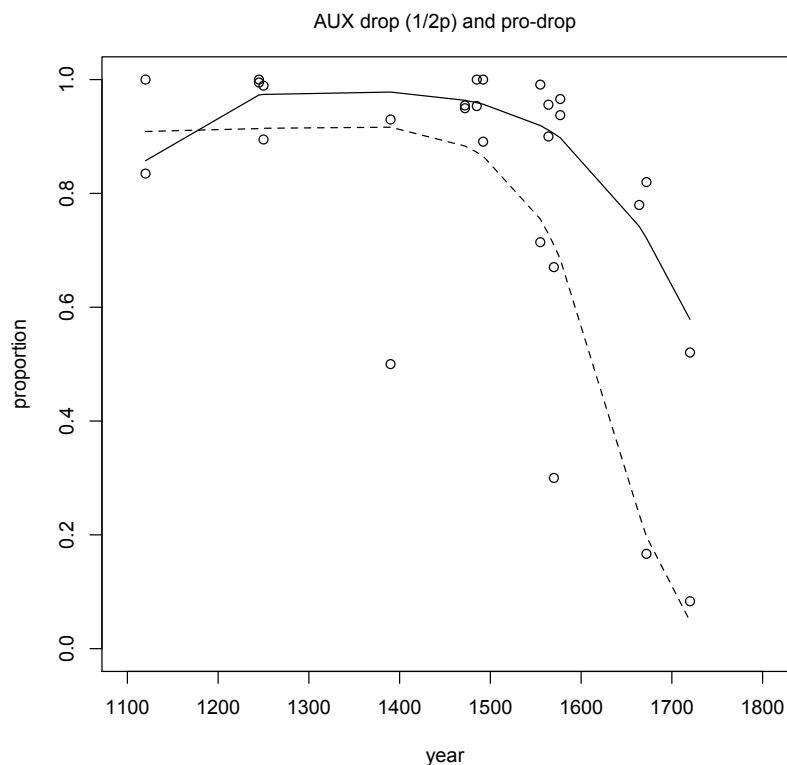
Figure 3.13: AUX-1/2P-drop (dotted line) and pro-drop overall (solid line), logistic regression

volves a non-null marker and does affect all tenses, so neither of the previous two points holds here. But then there is no *single* grammatical feature which could be impoverished, to begin with. Although an explicit analysis of OCS verbal paradigms in the relevant framework (i.e., Distributed Morphology) cannot be provided here, I conclude that the OCS and early ORu 2/3P.AOR/IMPF syncretism does not constitute counterevidence against Müller's (2006, 2007) revised version of the verb impoverishment hypothesis. However, it *is* problematic for morphologically superficial older accounts (like Jaeggli & Safir's 1989).[12]

---

[12]As discussed earlier, Müller (2006, 2007) argues that modern Ru is (still) a pro-drop language. Accordingly, he assumes that no impoverishment of person features is at stake in Ru past tense, because it never realised any person features to begin with – person features were part of the auxiliary specification and were historically dropped together with the latter. So then there is not an underspecified person morpheme, but no person morpheme at all, just as in Korean, Japanese and Chinese (Müller 2007, 8). Alternatively, Müller (2006, 14, fn. 20) also envisages the option (though refuting it at the same time) that Ru could have become a non-null subject language, and states that this could also be easily accomodated by his theory under the assumption that Ru past tense *was*, after all, impoverished. – Now, a potential problem is that the above points about Wambon carry over also to the Ru case: If any, then there is a null marker involved, and no parallel syncretism shows up in other tenses. So these same arguments could make an impoverishment account problematic. I have to leave this as an open issue here.

Since the period of the 16th-17th c. has proven so unstable, we now turn to the question of genre specificity and look at texts from more different genres from this time.

3.5.1.3.4 Samples by genre (16th/17th century)   The genres which we will compare here are *charters and court protocols*, written *correspondence* both in very elevated and in more colloquial style, and early *news* reports (*Vesti-kuranty*). These texts stem from a period of over 150 years, so we cannot quite exclude that diachronic change plays a role in addition to the differences in genre.

Charters (*gramoty***)**   The *gramoty* sample includes court protocols, testaments, land inspection protocols and donation acts from the early 16th century. The language of these charters is very formulaic, leading to a strong rise in the frequency of subject pronouns, as in the ubiquitous *sě azъ 'behold, I ...'*. Since these formulae typically introduce a topic at the beginning of a text passage, they have not been counted as familiar or background; so their higher frequency does not imply a change in the distribution of null vs. weak subjects.

Interestingly, the mere rise in token frequency of 1P pronouns – 185 instances among roughly 25000 tokens – does not affect 3P at all, for which we find exclusively pro-drop and no weak subject pronouns either in past or present tense. Thus, a sharp contrast arises in past tense, with PRT.1/2P pronominalisation (16:1 cases) vs. PRT.3P pro-drop (0:20 cases). Note that nominal subjects are by far most frequent, which is why numbers for pro-drop/weak pronouns are generally low. In present tense, pro-drop is still predominant in 1/2P, but with a lower amount than in any of the narrative samples discussed above.

	Aorist	Perfect	Past	Present
1/2P	–	1 (=100%)	–	19 (=82.61%)
3P	1 (=100%)	–	19 (=100%)	9 (=100%)

Table 3.17: Amount of pro-drop in court charters, 1480–1520 (n = 501; % of relevant subjects)

It is somewhat surprising that these gramoty show so little pro-drop, especially when compared with the findings by Borkovskij (1949) discussed above in section 3.5. There is a certain possiblity of a sampling error here; but a query for all 3P subject pronouns indeed yielded only 4 weak cases (Familiar Topics) out of 9 occurrences altogether. I conclude that early 16th century *gramoty* provide further evidence that 3P pro-drop was "lagging behind" 1/2P pro-drop diachronically.

Correspondence   Zaliznjak (2008) makes the interesting observation that the refined and stylistically elevated correspondence between the Duke Kurbskij and tsar Ivan the Terrible

is complemented by a very different style in the correspondence between Duke Grjaznyj and the tsar – apparently, the latter two entertained a less complicated relationship. As Zaliznjak shows, Kurbskij sticks to the traditional, periphrastic Perfect tense, whereas Grjaznyj favours the bare LPT, cf. tables 3.18-3.19 :

	Aorist	Perfect	Past	Present
1/2P	41 (=100%)	39 (=100%)	2 (=66.67%)	76 (=98.7%)
3P	3 (=100%)	–	8 (=80%)	9 (=75%)

Table 3.18: Amount of pro-drop in the *Letters of Duke Kurbskij to Ivan the Terrible*, 1564-1577 (n = 521; % of relevant subjects)

Indeed, the profile we observe in table 3.18 is almost the same as in early narrative samples, especially the chronicles: All preterite forms are well-attested, and so is pro-drop. The picture in the letters written by Grjaznyj differs, mainly with respect to the status of pronominalisation and the Perfect/Past divide:

	Aorist	Perfect	Past	Present
1/2P	–	7 (=77.78%)	2 (=9.52%)	29 (=82.86%)
3P	–	1 (=100%)	13 (=86.67%)	7 (=100%)

Table 3.19: Amount of pro-drop in the *Letters of Duke Grjaznyj to Ivan the Terrible*, 1570 (n = 348; % of relevant subjects)

Aorists have been dropped entirely here, and bare LPTs exceed the Perfect in frequency in 1/2P (21:9). Pro-drop in 1/2P has decreased considerably in Present and Perfect, and broken down completely in 1/2P.LPT. This is essentially the same picture as the one we found in the early 16th century *gramoty*.

Vesti-kuranty   The *Vesti-kuranty* from mid 17th century onwards – predecessors of first newspapers in Russian – consisted of translations of collected news from abroad written by semi-official, organised translators. Early influences from foreign syntax are expected to be found here. Occurrences of 1/2P are – unsurprisingly – very rare in this source. 3P pro-drop is reduced, but still clearly dominates in frequency over weak subject pronouns:

To the extent that this is on the right track, the *Vesti-kuranty* did not establish a non-pro-drop or partial pro-drop system in the written language, *despite the respective models* in several of the source languages from which they were translated (German, Dutch, French, English).

On the other hand, there is a very remarkable distribution of 3P pronouns in this source – the first 25 *vesti* – as a whole: Almost half of the 124 occurrences of subject *on* were classified

	Aorist	Perfect	Past	Present
1/2P	–	–	–	3 (=75%)
3P	–	–	16 (=66.67%)	27 (=87.1%)

Table 3.20: Amount of pro-drop in the *Vesti-kuranty*, 1656-1664 (n = 606; % of relevant subjects)

as Familiar Topics (60 (=48%) vs. 64 Shifting Topics). This is an extraordinarily high amount compared e.g. to the *Žitie* by Avvakum, which stems from the same period, but only contains about 10% weak 3P vs. 90% strong subject pronouns (see above). The influence of foreign models is very probable here. Weak 3P pronouns also occur in positions which are typical for a *partial* pro-drop language (cf. section 3.3.2), namely, as the coreferential subject of a complement clause:[13]

(67) Ис   Полши   пишѹт   что   они   велми   от   приходу   тѹрскихъ
      from Poland-GEN write-3PL that they-NOM very   from arrival-GEN Turkish

      людеи   боютца
      people-GEN   fear-3PL
      'From Poland they write that they very much fear the arrival of the Turks'

                                                                                  (Vesti, RRuDi)

Taken together, this finding indicates that the *Vesti-kuranty* did not constitute a radical departure from the widespread pro-drop properties of Ru of the time, but they did support the further dispersion of weak 3P subject pronouns by borrowing from the model of non-pro-drop languages. This type of selection from an autochtonous inventory of optional grammatical means conditioned by foreign influence is not at all exceptional.

3.5.1.3.5 Summary: Genre and pro-drop in 16/17th century   In a nutshell, this short overview of pro-drop in texts of different genres in 16th/17th c. indicates the following:

- Traditional texts which were strongly influenced by Church Slavonic, like the *Domostroj*, contain next to no weak subject pronouns, pro-drop being the rule.

- Similarly, pro-drop strongly prevails in the stylistically elevated correspondence of *Duke Kurbskij*, whereas it is much reduced in the more colloquial letters by *Duke Grjaznyj* (both 2nd half 16th c.).

- A corpus of court charters (early 16th c.), which have a high token frequency of subject pronouns mainly in formulaic contexts, does *not* show a decisive difference vis à vis

---

[13]On top of this, (67) shows an indefinite reading, to be discussed in chapter 5.

the high-variety literary texts, as far as the distribution of pro-drop and *weak* pronouns is concerned.

- In Avvakum's *Žitie* (late 17th c.), pro-drop vs. subject pronominalisation is used as a stylistic device; the mere frequency of subject pronouns is high, but they still show mainly strong uses.

- The *Vesti-kuranty* generally favour pro-drop over weak subject pronouns, but at the same time, they contain the highest *relative* amount of weak 3P pronoun uses in our corpus.

3.5.1.4 Conclusion

In the present section, referential pro-drop in Ru has been shown to have developed over time and diversified over genres. The diachronic development was not steady, but especially in aorist, present, and in the bare past in 3P, the tenses/persons with most evidence available, there was a clear decline in the amount of null subjects in relation to all weak pronouns and NPs, between the 12th and the 18th c. Two major outliers blur this picture: First, Nestor's chronicle has an unusual amount of weak subjects – for the most part, NPs which are mentioned repeatedly, and could just as well have been replaced by a null subject; but also weak subject pronouns, especially in speech reports. It seems that the chronicle indeed differs in its use of pro-drop from the strong Church Slavonic tradition surfacing e.g. in the Kiev paterikon. Second, the Domostroj has an unusually high amount of pro-drop for its time, which is obviously related to its large Church Slavonic parts, most of them taken over from earlier works in this tradition.

There is a clear rise in the amount of weak pronoun uses of all instances of the 3P pronoun *on* over time, i.e., *on* ceased to be an exclusively strong demonstrative; but this rise took place rather late, in the 17th and 18th c. samples. Use of 1/2P pronouns in bare past vs. Perfect in the early documents was dependent on the presence of an auxiliary. Examples of bare Past tense in 1/2 are very rare in the whole corpus, which is why the calculated amounts of null subjects vary so greatly. It is interesting to note that the amount of pro-drop in Perfect tense remains almost constantly (very) high: Wherever the past auxiliary is being used also in later texts, the correlation with null subjects is still strong.

The verb impoverishment hypothesis predicts that the changes in the past tense system affected the choice between null and overt subjects for the grammatical system as a whole. The closely parallel development documented in fig. 3.8 indeed supports this view. This holds more clearly in 1/2P than in 3P, which we attributed to the empty slot in the auxiliary

paradigm functioning as a meaningful morphological zero marker, as long as the evidence for a Perfect tense was still strong in the other persons (roughly between 13th and 16th c.). Assuming this morpheme, the verb impoverishment hypothesis is supported by our data.

The genre factor in the period of change (16th-17th c.) was scrutinised in the second part of the section. Correlations between the amount of pro-drop and the use of elevated style or register are obvious from the data. Received wisdom about the charters containing an especially high amount of weak pronouns, however, could not be confirmed: The pronouns are, for the most part, in a context which demands a strong reading. This is gradually different in the translated, vernacular *Vesti-kuranty*. The idea that an already existent development towards a partial pro-drop language may have been enforced by this type of language contact, gains in plausibility here.

3.5.2 Czech and Polish

3.5.2.1 Background

Both Cz and Pol count as canonical null subject languages today (with some qualification for Pol in Fehrmann and Junghanns 2008, see section 3.4.3), and do not appear to have changed much with respect to this property since their earliest stages.

For Cz, this seems to be the overall consensus in the literature (cf. Gebauer 2007; Trávníček 1961; Křížková 1974; Berger 1993 for pertinent remarks). But strong disagreement over the details has become apparent in the context of the history of Cz *expletive* subjects (for details cf. chapter 4).

For Pol, Pisarkowa (1984, 28ff) states that early texts, as e.g. the *Kazania gnieźnieńskie* (14th cent.), still realised all subjects overtly, whereas coreferential null subjects have occurred with a sizeable frequency only since the 15th cent., cf. the following example from the *Biblia staroszpatacka (= Biblia królowej Zofii)*:

(68)   Lepak rzecze   Bog: bądź   stworzenie w pośrzod wod,   a   Ø
       then   say-AOR.3SG god   be-IMP.3SG opening   in middle   water-GEN.PL and
       rozdzielił   wody.
       separated-PT.M.SG waters-ACC.PL
       'Then spoke God: There be an opening amidst the waters, and he split the waters.'

                                                              (BZ, 15th c., cit. Pisarkowa 1984, 28ff)

It would be surprising at least in two respects if this claim held true: First, the alleged change from a non-null subject language to a null subject language is typologically rare; we usually find the opposite direction of change (as in the case of Ru). Second, the alleged change

could not be related to verbal person marking, because no development towards less explicit marking is detectable in the history of Pol.

In sum, the apparent strict continuity from the oldest stages of Pol and Cz up to their current status as null subject languages is blurred by some problems: An increase in the amount of weak subject pronouns for Cz and a loss of the same for Pol have been reported in the literature. Whether these are just minor, accidental changes in frequency or more systematic changes affecting the status of Cz and/or Pol as null subject languages, is unclear. Furthermore, the origin and rise of overt expletive subjects can bear on this issue in Cz and – as we will see in chapter 4 – also in (middle) Pol.

The following remarks on pro drop in Old Cz are based on a detailed evaluation of queries for forms of *on* and *an* in the diachronic part of the Czech National Corpus (= 3640 hits before the year 1800) and the *Vokabulář Webový* corpus (= 3066 hits) (as of 2009). For Pol, all instances of *on(o)* and *an(o)* in our PolDi research corpus were analysed. Furthermore, a parallel corpus of bible translations was collected and evaluated for mismatches regarding the use of *on(o)* between diachronically distant versions of the text.

3.5.2.2 Czech

3.5.2.2.1 Pro drop and subtypes (homonyms) of *on*   Beyond any doubt, pro drop has been the rule in Old Cz since the earliest texts:

(69)   Když tak divně   Ø k nám přichodíš, Ø s   sobú   anjely
        when so   miraculous   to us   come-3SG   with self-INS.SG angel-ACC.PL
        přivodíš, [...]
        bring-2SG
        'When you come to us so miraculously, you bring angels with you'

                                        (ModlKunh, beginning 14th c., VokWeb)

(70)   I   vzkáza   mu   césařová, že Ø jej   chce   k tomu
        and tell-AOR.3SG him-DAT queen   that   him-ACC want-3SG to this-DAT
        připraviti, že   on sede   podlé žen   i   musí   tkáti. A   on
        prepare   that he sit-3SG next-to women-GEN.PL and must-3SG weave and he
        vzkáza   zase césařové, že Ø jí   chce   tak tkáti, že ona nikdy
        tell-AOR.3SG again empress-DAT that   her-DAT want-3SG so   weave that she   never
        nebude   sama tkáti, doniž Ø jest   živa.
        not-AUX-FUT.3SG self   weave while   be-3SG alive
        'And she had him be told that (she) wanted to prepare him that he sit next to the women and must weave. But he in turn had her be told that (he) wanted to weave for her thus that she would never weave herself while (she) was alive.'

                                        (PřKT, 1445, VokWeb)

(71) Tu   jich   Ø   počěchom prositi, aby   Ø   nechodili
     there they-ACC begin-AOR.1PL ask-INF   that-SBJ NEG-go-PT.PL to that
     k   tomu ješťerovi, ale   abychom Ø   šli   před   sě
     salamander-DAT but   that-SBJ-1PL go-PT.PL before SELF-ACC right-INS way-INS
     pravú cěstú.

    'Then we began to ask them not to go to this salamander, but that we take the right
    way in front of ourselves.'   (ŽsvO, 15th c., DČNK)

(72)   [A tak jich milostivě potresktav, v svuoj jě domek uvede. – 'And having punished
        them so gracefully, she lead them to her house.']

    i $Ø_i$ počě   jich pobiezěti, aby $Ø_j$ sebe   pokrmili. Tehdy oni_j
    and   begin-AOR.3SG them prompt-INF that-SBJ   SELF-DAT feed   then   they
    svaté skrúšenie vzemše,   tak sě $Ø_j$ rúče   polepšichu, že   […]
    holy   mercy-ACC take-CVB.PT so SELF-ACC   quickly improved   that
    '[And having punished them so gracefully, she lead them to her house.] […] and she
    began to prompt them to help themselves to something to eat. Having accepted the
    holy mercy, they improved so quickly that […]'   (ŽsvO, 15th c., DČNK)

A particularly early example of consequent pro drop is given in (69). (70) illustrates the
pattern of pro drop vs. subject pronominalisation which is typical of a null subject language:
If the subject of the (indicative) complement clause is coreferent with the main clause subject,
it gets dropped; if it is coreferent with another constituent, it is realised as a weak pronoun.
(71)-(72) shows furthermore that this pronominalisation is optional, as long as coreference
relations can be established on the basis of agreement information.

    As in modern Cz (cf. section 3.4), overt subject pronouns are used under emphasis/focus,
cf. (73):

(73)   [nemají v něm hřiechu za hřiech – 'they do not regard sin as sin']

    a coť svět   chválí, to i   oni.
    and what world-NOM praise-3SG that also they-NOM
    'and what the world praise, that they do also'   (ŠtŘb, end 14th c., DČNK)

and if the current Aboutness-Shift Topic changes, cf.

(74)   [Tehdy lvicě vytrhši sě jemu z rukú i poběže. – 'Then the lioness snatched away from
        his hands and ran.']

    A on také za   ní   poběže […]
    and he also behind her-INS ran
    'And he also ran after her.'   (ŽsvO, 15th c., DČNK)

(75)   [A tu sě jěchu toho kacieřě Manichea nutkati, aby šel v uoheň. – 'And then they started to force that Manicheus to go into the fire.']

Tehdy on, bojě sě ohně, nechtě jíti i počě sě
then   he   fear-CVB self-ACC fire-GEN NEG-want-AOR.3SG go and begin-AOR.3SG self-ACC
sápati
be-angry
'But he, being afraid of fire, did not want to go and began to be angry.'

(ŽsvO, 15th c., DČNK)

Specific syntactic conditions enforcing the use of a subject pronoun include the support of appositions, as well as correlative constructions like (76):

(76)   kdož to vidí, jest-li mocný anebo král, on přemůž
who   that see-3SG be-3SG-COND mighty or   king-NOM he defeat-IMP.3SG
své nepřátely
REFL-POSS-ACC.PL enemy-ACC.PL
'who sees this, if he is mighty or a king, may he defeat his enemies'

(Snář, 1581, DČNK)

The "test candidate" for a weak pronoun, consequently, is a Familiar Topic with a close subject antecedent. Some such cases indeed occur in early texts, cf.

(77)   velicí-li zubové u slona čili malí oni
big-QU   tooth-NOM.PL at elephant or   little-NOM.PL they
'whether an elephant's are big or they are small'   (ŠtŘb, end 14th c., DČNK)

(78)   [Ten jistý césař Octavianus Augustus ... zkazil mnoho panen. – 'This emperor O. A. ... ruined many maidens.']

On jest vyhnal svú vlastní ženu [...]
he   AUX-3P chase-off his-ACC own   wife-ACC
'He chased out his own wife [...]'   (PřKT, 1445, VokWeb)

More commonly, we find weak 1/2P subject pronouns:

(79)   A také, když já stojím, vie Buoh, že já stojím. A když sědu,
and also when I stand knows God that I stand and when sit-1SG
vie, že nestojím.
know-3SG that NEG-stand-1SG
'And also, when I stand, God knows that I stand. And when (I) sit down, (he) knows that (I) do not stand.'   (HvKrJ, mid 15th c., VokWeb)

The second sentence of (79), which contains three instances of pro drop, highlights the fact that what we are facing here is not an all-or-nothing situation, but rather *optionality*, an alternation of null subjects with (much rarer) occurrences of weak subject pronouns. The central question for the present section is thus: To what extent did the relative frequencies of the various subtypes of *on(en)* change over time? Or more specifically, did the amount of weak subject pronoun uses of *on* expand or reduce diachronically?

As an aside, it must also be noted that entities which are so prominent as to always facilitate coreference – typically God or Jesus in religious texts – can attract coreference with *on*. This group of examples has to be kept seperate as belonging to deictic rather than to anaphoric reference, and does not bear on the issue of pro drop in general.

(80) Ale knieže to, což duostojného jest kniežěte, mysliti bude a
      but duke   that which worthy-GEN.PL be-3SG duke-GEN.SG think be-FUT.3SG and
      on nad vévody stane.
      he above rulers stand-3SG
      'But the duke will think that which is worthy of a duke, and HE (= God) stands above the rulers.'   (PřIzJD, end 14th c., DČNK)

Further types (homonyms) of *on*, which have to be kept apart from the weak 3P personal pronoun, are the demonstrative *on(en)* and the expletive *on*. The *demonstrative* can be syntactically attributive or independent, just like *onen* in modern Cz; morphologically, its M.SG.NOM/ACC. can take the shape of either *onen* or *on* (cf. (81)) in 14th cent. texts, but becomes only *onen* by the 16th cent. All other inflectional forms are identical with those of the present-day personal pronoun *on/-a/-o*.

(81) a. A král ... sěka on ľud hlúpý
          and king-NOM   cut-CVB that-ACC.SG people-ACC stupid
          'And the king ... cutting that stupid people.'   (AlxB, ca. 1350, VokWeb)

      b. ženúc ot krále onoho /[...] s člověkem [...] jehož
          chasing from king-GEN that-GEN.M.SG   with person-INS   who-ACC
          on král vypovědě
          that-NOM.SG king-NOM expel-AOR.3SG
          'chasing away from that king ... with a person ... whom that king had expelled'
                                                        (AlxB, ca. 1350, VokWeb)

      c. kdež sě on ľud kryjě lazí
          where REFL on-NOM people hide-CVB.PRS creep-3SG
          'where this people, hiding, creeps'   (AlxM, ca. 1350, VokWeb)

Expletive on will be discussed in some more detail below. Its most unusual feature – beyond its mere existence in so clearly a pro drop language – is its doubling a lexical subject, with

which it obligatorily agrees:

(82) On jistý Krok   zpravoval jest   zemi,   [...]
     on certain K.-nom rule-pt.sg.m. aux-3sg earth-acc.sg
     'A certain Krok ruled the world, [...]'   (PulkKr, ca. 1400, DČNK)

Due to this behaviour, it is sometimes difficult to decide between an analysis of *on/-a/-o* as an attributive demonstrative and as an expletive. (81-a) is attributive beyond doubt, with *on* in the accusative; in (81-b), an attributive interpretation is more plausible than an expletive one; in (81-c), the attributive analysis seems semantically more adequate, and *on* as an expletive would occupy an unusually low position.

3.5.2.2.2 Parallel (bible) corpus study In order to get an overview of potential changes in the uses of *on* between an Old Cz and a corresponding modern Cz text, I conducted a small study of a parallel bible corpus which contains three translations aligned at verse level: the *Bible kralická* from 1613 and its immediate predecessor, the *Biblí šestidílná* (1579–1593) (only New Testament), as well as the recently published *Nová kralická bible* (2009). The electronic texts were prepared by Michal Rovnaník (cf. http://www.etf.cuni.cz/~rovnanim/bible/index.html). I aligned them and transferred them into a searchable multilingual New Testament database.

In the majority of cases in which *on* appears, it does so in all three versions of the text. 152 of the 564 instances of *on* in the modern text *only* appear there (= the "exclusively new" instances), and 196 of the 572 instances in the *Bible kralická* from 1613 do not appear in the modern version(= the "exclusively old" instances). The "exclusively new" group typically contains cases of topic shift (83-a) or beginnings of verses; the latter often license pro drop by coreference with the most salient characters (Jesus), which is dispreferred in modern Cz (BiblKral: *I Ø řekl: ... 'and said'* vs. NBK: *A on řekl: ... 'and he said'*). More rarely, the modern personal prononun corresponds to a demonstrative or a relative pronoun in the old translation (84)-(85-a):

(83) a. ... dím   tomuto: Jdi,   a jde,   a jinému: Přiď,   a
              say-1sg that-one-dat go-imp.2sg and go-3sg and other-dat come-imp.2sg and
          přijde, a   služebníku svému: Učiň   toto,   a učiní.
          come-3sg and servant-dat refl-poss.dat do-imp.2sg that-acc and do-3sg
          'I tell one of them "Go", and he goes, and the other "Come", and he comes; my
          servant, I tell "Do that", and (he) does it."   (Lk 7:8, BiblŠest 1579)
      b. Tomuto   řeknu 'Jdi'   a   on   jde   a   jinému: 'Přijď'
          that-one-dat say-1sg go-imp.2sg and he-nom go-3sg and other-dat come-imp.2sg

a   *on*   přijde; svému   otroku   řeknu 'Udělej tohle' a
and he-NOM come-3SG REFL-POSS.DAT servant-DAT say-1SG do-IMP.2SG that-ACC and
udělá to.
do-3SG that-ACC

(NovBiblKral 2009)

(84) a. *Kteráž<to>* řekla: Žádný,   Pane.
which-F.SG say-PT.F.SG noone-M.NOM.SG master-VOC
'who said: Noone, master.'   (BiblKral 1613)

b. *Ona* řekla: ``Nikdo,   Pane.''
she   say-PT.F.SG noone-M.NOM.SG master-VOC

(NovBiblKral 2009)

(85) a. Mistře,   kdo zhřešil, *tento*-li, čili rodičové jeho
master-VOC.SG who failed   that-NOM-QU or parent-NOM.PL REFL-POSS
'Rabbi, who failed – he or his parents'   (BiblKral 1613)

b. Rabbi, kdo zhřešil -- *on*,   nebo jeho   rodiče
master   who failed   he-NOM or   REFL-POSS parent-NOM.PL

(NovBiblKral 2009)

The "exclusively old" examples, on the other hand, rather concern construction-specific occurrences of *on*, such as with temporal or modal *když- 'when/if'-* adverbial clauses, or after verbs of perception, cf.

(86) a. I přišed   k nim,   nalezl   je,   a oni   zase spí
and come-CVB to they-DAT find-PT.M.SG they-ACC and they-NOM again sleep-3PL
'And when he came to them, he found them, and/as they were again sleeping.'

(BiblŠest 1579)

b. A když přišel,   znovu je   nalezl,   jak spí
and when come-PT.M.SG again   they-ACC find-PT.M.SG how sleep-3PL

(NovBiblKral 2009)

In sum, we find slight changes in distribution, but no general rise or decline of usage as a weak pronoun over time. Cz behaves like a canonical null subject language both in synchrony and in diachrony.

3.5.2.3 Polish

3.5.2.3.1 From pronoun to pro drop?   As mentioned above, Pisarkowa (1984, 28ff) in essence claims that the frequency of referential null subjects in Pol increased diachronically. She makes explicit reference to the *Kazania Gnieźnieńskie* for evidence of early obligatory subject pronominalisation. The *Kazania Gnieźnieńskie* (1st half 15th cent.) indeed stick out by their incredibly consequent use of pronominal subjects. In our test sample of about 5300

tokens, only one (!) instance of referential pro drop vs. hundreds of weak subject *on/-a/-o* could be identified. Even more interestingly, the *Kazania Gnieźnieńskie* often use them in the manner of clitic-doubled subjects, cf.

(87) Czwarteć jest   on [w] pościeli {tako} ubo[g]i był,   iżeć on
      fourth   AUX-3SG he in bed   so   poor-NOM be-PT.M.SG that he
      namniejsze pierzynki   nie miał{ci}   jest {je}   on
      smallest-ACC.PL blanket-ACC.PL not have-PT.M.SG-you-DAT AUX-3SG they-ACC.PL he
      był   {w ten to czas, gdyżci się jest   on był   narodził},
      be-PT.M.SG in that-ACC time-ACC when REFL AUX-3SG he be-PT.M.SG born-PT.M.SG
      aliżci   jego {miła} matuchna {Maryja} trochę siana   jest{ci
      but-MP-DAT his dear   mother   M.   little-ACC hay-GEN AUX-3SG-you-DAT
      ona} była   [w] jasły   podeń   podłożyła.
      she be-PT.F.SG in cot-ACC.PL under-him put-under-PT.F.SG
      'Fourth, he was so poor in his bed that he did not have the smallest blanket at the
      time when he was born, but his dear mother Mary put a little hay under him in the
      cot.'   (KazGn, PolDi)

However, Pisarkowa's conclusion is premature: Comparable, even earlier texts like the *Kazania Świętokrzyskie* (mid 14th cent.) contain a large amount of null subjects, cf.

(88) widziech, prawi, anjeła   bożego   mocnego   s   nieb(a sle)ciew,
      see-AOR.1SG say-3SG angel-ACC Godly-ACC.SG mighty-ACC.SG from heaven fly-CVB
      prawi, postawi   prawą noge   na morzy   a   le(wą na ziemi)
      say-3SG set-AOR.3SG right foot-ACC on sea-PRP.SG and left-ACC on earth-PRP.SG
      '(I) saw, (he) said, God's mighty angel flying down from heaven, (he) says, (he) set
      his right foot on the sea and his left foot on land'   (KazŚw, PolDi)

In fact, in our about 3500 tokens sample of *Kazania Świętokrzyskie*, weak pronominal subjects are outnumbered by *pro* by a factor of 12 (6 weak subject pronouns vs. 72 x pro drop). We suppose that the fact that the *Kazania Gnieźnieńskie* constitute such an outlier must be motivated dialectally.

3.5.2.3.2 Parallel (bible) corpus study As in the Cz case, a parallel corpus of different bible translations was consulted in order to get an impression of major changes in the realm of pro drop. It consisted of the New Testament from the *Biblia Brzeska* (1563), the *Biblia Gdańska* (1632) and the *Biblia Poznańska* (1991). As above, electronic versions of the texts were available from concordance sites on the internet, and were aligned by verse and entered into a searchable database.

Cases in which *only the modern text* (*Biblia Poznańska*) contains a subject pronoun come

in three major subgroups: First, the demonstrative lexeme may have changed (e.g. *ten*, *oto* or *tenci* in earlier variants vs. *on* today – (89)). Second – as in Cz above – the old variants often use a null subject coreferent with the extremely salient protagonist Jesus; this is replaced by *on* today. Third, earlier examples sometimes allow for topic shifts without subject pronominalisation, whereas this is necessary today – (90).

(89)   a.   *On*   wyrzuca   czarty   mocą   Beelzebula
            that-one throw-out-3SG devil-ACC.PL power-INS   B.-GEN
            'He throws out devils with the might of Beelzebul.'   (BiblPozn)

       b.   *Ten*   nie wygania diabłów   jedno przez Beelzebuła
            that-one not dispel-3SG devil-ACC.PL only   by   B.-ACC
                                                                  (BiblBrz, ParBib)

(90)   a.   I   zgromadzą się   przed Nim   narody   ..., a *On* oddzieli
            and gather-3PL REFL before he-INS nation-NOM.PL   and he   separate-3SG

            jedne   od   drugich
            one-ACC.PL from other-GEN.PL
            'And the nations gather before him ... and he separates them from one another.'

            (BiblPozn)

       b.   I   będą   zgromadzone   przedeń   wszytki   narody   i
            and aux-3PL gather-CVB.PT.PAS-PL before-him   all-NOM.PL nation-NOM.PL and

            odłączy   jedny   od   drugich
            separate-3SG one-ACC.PL from other-GEN.PL
                                                                  (BiblBrz, ParBib)

While the latter point may be taken as evidence for an increase in weak pronominal subject use, it is also important to note that even a clearcut canonical null subject language like Italian would require a pronoun in cases of topic shift (Frascarelli 2007). In this sense, modern Pol is just as good a null subject language as Old Pol used to be. Moreover, in some cases of topic shift, it is only the Old Pol *Biblia Brzeska* (1563) that contains a subject pronoun:

(91)   a.   Wyciągni   rękę   twoję.   A   *on* wyciągnął   i
            outstretch-IMP.2SG hand-ACC.SG your-ACC.SG and he   outstretch-PT.M.SG and

            stała   się   zdrowa   (BiblBrz, ParBib)
            become-PT.F.SG REFL healthy-NOM.SG
            'Stretch out your hand. And he stretched it out, and it became well again.'

       b.   Wyciągnij   twoją   rękę.   I   wyciągnął   ją.
            outstretch-IMP.2SG your-ACC   hand-ACC   and outstretch-PT.M.SG it-ACC
            (BiblPozn, ParBib)

An important further case of (attributive) pronoun use in the Old Pol version vs. null

realisation in the modern one concerns *attributive on*, as in

(92) a. A *on* sługa upadszy kłaniał mu się mówiąc
          and that-NOM.SG servant-NOM.SG fall-CVB.PT bow-PT.M.SG he-DAT REFL say-CVB.PRS
          ... (BiblBrz, ParBib)

          'And the servant that had fallen to his feet, bowed down before him and said ...'

     b. A sługa upadłszy do nóg oddał mu pokłon,
         and servant-NOM.SG fall-CVB.PT to feet-GEN give-PT.M.SG he-DAT bow-ACC
         mówiąc ... (BiblPozn, ParBib)
         say-CVB.PRS

The latter example illustrates the little known fact that Old Pol still had an attributive demonstrative, homonymous with the personal pronoun of today, which was later replaced by *ów 'that'*. Actually, (92) could also support an alternative analysis with an agreeing, subject doubling expletive *on,* akin to the Cz expletive. We will return to this point in the next chapter.

In the present subsection, we have seen evidence that the null subject property of Pol changed only slightly in diachrony. The counterexample of *Kazania Gnieźnieńskie* is a stunning outlier, but pro drop could be firmly attested in other early sources.

3.5.2.4 Conclusion

The upshot from our research on Cz and Pol weak pronouns as well as the parallel bible corpus is obvious: Both languages underwent minor variations in the use of weak pronouns vs. null subjects, but no clearly directed change could be detected. An important interfering factor is the use of *on, -a, -o* as demonstrative pronoun and as (a specific kind of) expletive. The alleged widespread use of weak subject pronouns in early Pol (Pisarkowa 1984) could be shown to be an unusual outlier in one important text. The precise reasons for this unusual behaviour (dialect, register etc.) would be a rewarding topic for further research. Another question to be clarified is how exactly the rather frequent and unproblematic use of weak 3P subject pronouns in modern Pol came about. I suspect that this must be a later development outside the time frame of our study.

3.6 Summary

In the present chapter, we first went through the major general linguistic approaches to the null subject phenomenon available today. Two leading ideas were singled out to be

confronted with the diachronic data of Ru, Pol and Cz: (i) the *grammar-oriented* verb impoverishment hypothesis, and (ii) the *context-oriented* referential and information-structural conditionality of the use of weak subject pronouns. Both hypotheses have been applied extensively in synchronic studies of referential null subjects in Slavonic, which were discussed together with their predictions for diachrony, in the second part of the chapter. In the third part, we turned to studies on referential null subjects in the historical domain, and to our own corpus studies. The main result of these is a confirmation for the verb impoverishment hypothesis in the history of *Ru*, in two respects:

At the level of language use/individual examples, there is a strong inverse correlation between an overt auxiliary in the preterite tenses and the use of a 1/2P pronoun, which holds for a very long time, up until the 18th c. In 3P, pro-drop remains frequent for some time despite the decrease of explicit agreement marking on preterite verbs. Apparently, the meaningful gap in the auxiliary paradigm was still informative; in any event, use of 3P pronouns did not follow the same conditions as 1/2P. As far as our data are concerned, the rise of the weak subject *on* began in 17th c. and may have been supported somewhat by translations from West European sources (*Vesti-kuranty*); however, the source languages of the latter were all but homogeneously non-null subject languages (Polish!). – At the level of its impact on diachronic change, the verb impoverishment hypothesis is a lot more general: A small-scale change in the morphological derivation – not just any overt syncretism – conditions a major change in pro-drop properties across tenses and persons. The parallel quantative development between both phenomena, which could be shown here for Ru, supports this view.

Furthermore, I argue that conceptually, the context-oriented conditions of pronoun use must be distinguished more effectively than in earlier accounts, in order for the quantitative data to make any sense. When this is being done, received opinions, e.g. about the small amount of null subjects in charters of the 16th/17th c. – as alleged examples of the "real" East Slavonic vernacular – often lose empirical support.

Cz and Pol, despite their both being canonical null subject languages, show some synchronic differences in the use of subject pronouns (as discussed in the second part of the chapter). Diachronically, we did find some mismatches between the conditions on the use of null subjects vs. pronouns in older vs. modern versions of the same text (in a parallel bible corpus), but there was no clear evidence for a development in one or the other direction (*contra* claims in the literature on Pol). The use of weak pronouns remained rare throughout, except for some "outliers" like the early *Kazania Gnieźnieńskie*. This leaves open the issue of how the fine-grained synchronic differences between the two languages came about. I assume that they may have developed later on, in 18th-19th c., when further important changes

occurred in the use of indefinite impersonals in Pol. Furthermore, they may be related to genre and register, as is well-known in the case of Cz.

An interesting phenomenon often intervening with the search for weak pronominal subjects in Cz and Pol diachrony, are apparent *subject expletives*. As noted in the theoretical part of the present chapter, subject expletives are an unexpected phenomenon in a canonical null subject language, which has often raised curiosity about their origin and current status. The discussion centers around Cz, but some important historical evidence from Pol also bears on the matter, and is often neglected. Subject expletives are the topic to which we turn in the following chapter.

4 The History of Expletives in Czech and Polish

4.1 Introduction

Across the boundaries of linguistic theories, it has been assumed for long that null subject languages cannot have subject expletives. The correlation has intuitively appeal – after all, the syntax of null subject languages does not forge the use of an overt referential subject, so why should it enforces an expletive? Haider (2001, 285) cites Gilligan (1987) and Auwera (1984) with the finding that of all the properties usually associated with *pro* drop in generative grammar (cf. Rizzi 1982; 1986), the lack of expletive subjects is the only one which actually holds without exception cross-linguistically (cf. also the discussion in Newmeyer 2004). The same factual claim is implicit in grammaticalisation approaches: Lehmann (1995) introduces the following grammaticalization path:

(1)   demonstratives > personal pronouns > expletive subjects > subject clitics > verbal affixes

The end product implies null subjects, which should thus be in complementary distribution with expletives (if developed along this path). According to the metastudy of Diessel (1999), demonstrative pronouns form only *one* possible source of expletives, so in principle, demonstrative pronouns or the products of their further grammaticalisation (verbal affixes, null subjects) might co-occur with (a different type) of expletives. But the type of expletives which we are facing in the case of Cz is just the one which should be derived along the path in (1).

Contrary to these claims, Holmberg (2005) has lately shown that Finnish, a *partial* null subject language, at least optionally realizes subject expletives. Cz, as discussed by Lindseth (1998), could be a further candidate for a counterexample: We saw in section 3.4 that it carries all hallmarks of a canonical null subject language . Nevertheless, it has optional expletive-like elements, namely, agreeing *on, -a, -o* (cf. (2)), and non-agreeing *to*. I will only deal with *on, -a, -o* and its relation to to the weak pronoun in the remainder.

(2)   a.   Ono   pršelo.   [Cz]
               it-N.SG.NOM   rain-PT.N.SG
               'It rained (a lot).'

        b.   Ona   se   matka   zlobila.   [Cz]
               she-F.SG.NOM   REFL.ACC   mother-F.SG.NOM   be-angry-PT.F.SG
               'Mom was rather angry.'   (Lindseth 1998)

What is less well-known is that Pol also had an expletive *on / an*, at least historically. Present-day Pol uses *to* or *oto* in partially overlapping functions:

(3)   Oto my, książęta   Kościoła,   liczymy pieniądze   Kościoła   ... [Pol]
       OTO we priest-NOM.PL church-GEN.SG count-3PL money-ACC.PL church-GEN.SG
       'Behold (Cz. *hle*), we, the priests of the church, count the money of the church'

(Intercorp)

4.2 Synchronic properties of Cz expletive *on*

Lindseth (1998) points out that Cz expletives of the type illustrated in (2) are exceptional in several respects. On her view, the key observation is that Czech expletives actually do convey meaning, namely, introduce special *emphasis* or at least have a stylistic function. Thus, they are not to be taken as parallel to *weak* thematic (= non-expletive) pronouns, but to emphatic, strong thematic pronouns. Such strong pronouns are to be expected also in a canonical null subject language. In generative syntactic terms, they do not fall under the rule of Chomsky's (1981) Avoid Pronoun generalization. The emotional, emphatic overtone is attributed (by Lindseth and others, see below) to the *diachronic* source of the Czech expletives, following Trávníček (1961) and, ultimately, Zubatý (1909): They are traced back to a deictic particle *ano*, corresponding to Latin *ecce 'look!'*.

Lindseth (1998, 158/162) further observes that Cz expletives can generally occur in embedded clauses, but only a subclass of them are acceptable in *wh* questions, namely those which correpond to a thematic subject:

(4)   a.   *kdy   vono prší   / prší   vono
             when it-N.SG rains-N.SG   rains-N.SG it-N.SG

       b.   *proč vono se   tu   včera   nepracovalo
             when it-N.SG REFL here yesterday not-worked-N.3SG

(Lindseth 1998, 159 (23)/(24))

(5)   a.   kdy   vona se   tu   naskytla   ta   vosoba
             when she   REFL here appeared-F.SG that-F.SG person-F.SG
             'When did that person appear here?'

       b.   proč von mu   otec   všechno dovolí
             why   he-SG him-DAT father everything allows-PRS.3SG.PF
             'Why does (his) father allow him everything?'   (Lindseth 1998, 162 (36)/(37))

Three relevant subtypes of expletive constructions in Czech can be established, according to Lindseth: (i) With meteorological verbs and non-agreeing (impersonal) periphrastic passives, i.e., cases of expletive *pro*, overt expletive subjects encode surprise on behalf of the speaker at the propositional content, and concomitant questioning of the content. This addi-

tional surprise/questioning is infelicitous in a wh question for pragmatic reasons (4), Lindseth reckons. (ii) In subject doubling cases, expletives mainly indicate a lower, colloquial register, and need not involve emotive emphasis; consequently, they allow for questioning, as in (5). (iii) Expletive subjects as correlates of extraposed clauses are optional in Czech. Lindseth argues that the existence of this type of expletives is, generally, independent of the setting of the pro drop parameter. She speculates that Cz *ono*-correlates were derived from the original demonstrative pronouns independently (Lindseth~1998,~177). The lack of an obligatory emphatic, emotive function of Cz expletives in cases (ii) and (iii) had been pointed out before by Hirschová (1984).

The "doubling" of the subject by the expletive, as in (5), as well as the obligatory agreement of the expletive with its "associate" are the two major challenges to any theory of Cz expletives. Synchronic generative treatments have unanimously argued that expletive *on, -a, -o* occupies the usual overt subject position, i.e., the specifier of TP (Rezac 2004) or, equivalently, the specifier of IP (Lindseth 1998, 163). The agreeing associate (subject) NP is then situated in a lower argument position, namely, Spec-VP.

Rezac (2004) describes the syntactic properties of Cz *on, -a, -o* within a generative theory of expletives and similar elements in Romance and Germanic. Cz expletives are non-clitic pronominals, prosodically fully integrated into their clause, immediately follow the C projection and precede background material. A nice demonstration of the lower bound of the expletive position is the fact that if the verb precedes *(v)on,-a,-o*, the sentence must be interpreted as an interrogative, which is typical of verb fronting to a high (C^0) position:

(6) Viděl   von tam Honza někoho
     saw-PT.M.SG he   there H.-NOM someone-ACC
     'Did John see someone there?/*[What happened?] John saw someone there.'

(Rezac 2004, 243 (299))

According to Rezac (2004, 239), there is no restriction whatsoever on the semantic type or information-structural status of the associated NP, which may be a quantifier, a definite or (specific or non-specific) indefinite, or even an idiomatic non-referential DP, and contrastive or non-contrastive topic, new-information focus or focus bound by a operator like *only* (respective examples are deferred to an unpublished paper inaccessible to me [RM]). It only has to be in 3rd person. The expletive agrees with AGR (the verbal agreement category), not necessarily directly with the features visible on the associate, cf.

(7) a. *Vona/voni   tam   přišli   [Dana   a   Radek] pozdě.
         she/they-NOM there came-3PL   D.-F.NOM and R.-M.NOM late

b. Vona/*voni tam přišla   [Dana   a   Radek] pozdě.
    she/they-NOM there came-F.3SG D.-F.NOM and R.-M.NOM late
    'Dana and Radek came there late.'   (Rezac 2004, 242 (296))

In a similar vein, when there is no associate, as in impersonals, the expletive appears in the default 3rd neuter singular form realized also on AGR.

Syntactically, Rezac's central observation is that the expletive may never be *preceded* by its associate, i.e., crossed over by the latter when it moves into the CP domain. In this respect, Cz expletives sharply differ from non-agreeing French *il* (8-b) or English *there* (8-c):

(8) a. Kolik   koláčů   (*vono) zbylo?
         how-many cakes-M.GEN.PL it-N.3SG remain-PT.N.SG

     b. Combien de gâteaux il restait?
         how-many of cakes   it remained
         'How many cakes did there remain?'   (Rezac 2004, 247 (307))

It also differs, in the same respect, from subject clitic doubling in dialectal Italian (Fiorentino) and in Spanish (Rezac 2004, 248f), which pattern with French and English (8-b,c). Generally speaking, such a crossing restriction seems to be specific to agreeing, phrasal expletives, as they are present in Cz. It is the overtly visible agreement features on the expletive which make it count as a pronoun in the sense of binding theory and lead to the crossing restriction (Rezac 2004, 254).

Furthermore, he assumes (p. 263) some link to the constituency of the clitic and its associate: In cases like (8-b,c) they may be originating from the same DP constituent. Conversely, the Czech expletive and its associate cannot form a discontinuous constituent, and the expletive must be generated in some non-thematic position. As it stands, this "parity of reasoning" (p. 264) argument seems quite unconvincing to me. The binding violation would be sufficient to exclude cases like (8-a).

It is a tricky question how to model the relation between the expletive and its associate in syntax. Obviously, it cannot be a binding relation in the technical sense, because pronouns have to be free in their binding domain (condition B) and referential expressions have to be free everywhere (condition C of Chomsky's 1981 binding theory and its successors). More-over, *all of* (7) would end up as a configuration like (8-a), which is strongly ungrammatical. Rezac (2004, 258 (321)) formulates the generalization that an expletive *can* act as a binder (in the sense of binding theory) for a DP α only if the expletive enters into an agreement relation with a different DP β, $\alpha \neq \beta$ (via the AGR head). Thus, non-agreeing expletives cannot bind anything, since they do not agree with any DP; and agreeing expletives can only bind a DP

distinct from their associate. Applying these ideas to Czech, we find that the associate is itself the source of the φ-features on T (AGR), with which the expletive agrees; therefore, the associate must not be anaphorically bound by the expletive, but it may be a pronoun.

Concerning the overt agreement of the expletive, Rezac proposes to relegate these effects to the lexicon. Both French *il* (9) and English *there* are restricted to unaccusative structures and indefinite, 3rd person associates, but they differ in number agreement:

(9)   Il est/ *sont arrivé   trois linguistes.
      it is   are   arrived-sg three linguists
      'There have arrived three linguists.'   (Rezac 2004, 267 (328))

Since in presentational constructions with expletive *ce* in French, number agreement with the postverbal subject is at least an option, the restriction in (9) must be due to properties of the expletive *il* itself, i.e., it is lexically specified for singular number. The Czech expletive varies in number and gender, but it is always 3P; on the other hand, the Finnish colloquial split subject construction reported in Holmberg and Nikanne (2002) allows for 1P or 2P associates. So variation in person agreement is not generally precluded, and the Czech expletive must be specified for 3P in the lexicon.

Summing up, the most interesting synchronic properties of Cz expletives, which also drove the approaches by Lindseth (1998) and Rezac (2004), are:

- doubling of a thematic associate subject

- agreement with the associate

- emphasis on the pronoun, depending on the exact subtype of

- the crossing restriction – the associate must not precede the expletive in overt syntax

4.3 Diachronic properties of Cz expletive *on*

4.3.1 Previous approaches

The history of the Czech expletive *on,-a,-o* has been documented and discussed in quite some detail in the traditional literature. Zubatý (1909) was the first to argue that the Czech *expletives on,-a,-o* were genetically related to a non-agreeing, emotive, demonstrative interjection *ano*, rather than to 3P personal pronouns. Trávníček, who outright euphorically acclaimed Zubatý's findings (*podle mého mínění největší jazykozpytný objev v novější době* – "on my view,

the greatest linguistic discovery in recent years" – Trávníček 1930, 10; *veliký, ba epochální význam* – "*great, even epoch-making importance"* – ibid., 58), extended this line of thought into a larger theory of the historical impact of interjections on clausal syntax. The reasons for Trávníček's enthusiasm seem to be that, firstly, Zubatý's idea does not presuppose the existence of weak personal pronouns in early Old Czech (which are not attested), thus works better than competing proposals. Second, it can explain away the "doubling" of the thematic subject, which becomes only apparent, once the alleged expletive is analysed as a particle. Thirdly, the proposal provides a historical source for the emotive overtones of modern Czech expletives (see above) – they are derived from emphatic demonstrative particles. This presupposes, of course, that such semantic/pragmatic nuances could be preserved in the lexicon and grammar for many centuries.

Zubatý states that the Czech expletive *on, -a, -o* was derived from a non-agreeing, emotive interjection *ono*, termed *náladové ono "emotive ono"* by Trávníček (1930). According to him, Jungmann's description saying that "*on ante nomen ponitur claritatis gratia"* ("*on is put before the noun [i.e., the subject] for clarity's sake")* was misleading, because it overlooked the emotive overtones of the expletive, which stem from the original interjection. The historical development from *ono* to agreeing *on,-a,-o* – following Zubatý – ran in a fashion parallel to the development of earlier *ano* into *an, -a, -o.* Morphologically, *ano* is a contraction of the coordinative conjunction *a* with (uninflected) *ono* from preliterary times, which later became separated again.[1]

Let us first focus on non-agreeing *ano* and its inflected counterpart *an.*

4.3.1.1 *ano* **and** *an*

Zubatý (1909), but much more strongly Trávníček (1930, 120ff), take issue with the harmless hints at an analysis implicit in Gebauer's 1903 dictionary entry for *an* (Gebauer 1970): *An* is a contraction of *a+on*, originally a coordinating conjunction meaning sometimes *'and he'*, sometimes *'but he'*, later also understood as a subordinating relativiser (lat. *qui*). In the syntax part of his *Historická mluvnice jazyka českého*, edited by Trávníček himself in 1929 (Gebauer 2007), Gebauer cites *ano* in line with *a já, a ty* etc. and mentions that *ano* acquired a subordinative meaning (*když 'when', že 'that'*) which could later also be expressed by the masculine *an* without agreement (p. 677). Gebauer does not seem to view *on, -a, -o* as a kind of particle; he calls its expletive use *předjímání podmětu* (p. 218) *'anticipation of the subject'*,

---

[1]Zubatý and Trávníček, of course, note that *a+on* should have led to *án* by the Old Czech rule of vowel lengthening, but decide to treat this as an exception. They similarly gloss over the curious fact that a fixed cliticization is supposed to become reverted.

similar to the sort of "anticipation" of the object we find in *tobě mému bratrovi 'to you, my brother'*, i.e., essentially, an apposition structure – an interesting idea by itself, which was never explicated further (but see below).

Ano and *an* are interesting objects of study all by themselves. Syntactically, *ano* developed in several historically documentable phases from an emotive particle first (i) into a subjunction for perception verb complements, later (ii) into an adverbial clause subordinator and relativiser, and finally, only since the 19th century, (iii) into the modern answering particle *('yes')*. Regarding the rise of the agreeing forms *an,-a,-o*, Zubatý and, in more generality, Trávníček, argue that agreement is a consequence of the former particle losing its autonomy and becoming integrated into the following clause. During a transition period – around the second half of the 14th c. –, non-agreeing *ano* and *(a) ono* were possible with a MASC.SG or FEM.SG/PL subject as well; the same may still hold for dialects (see below).

Concerning *an(o) / a+on(o)* in subordinate clauses, Zubatý (1909) observes an interesting distributional restriction which remains essentially valid throughout Old Cz: If the embedded clause contains an overt subject, only *ano* is used (10-a). If the embedded clause contains no overt subject, but its subject is coreferent with an antecedent in the main clause, *an* is used (10-b):

(10)   a.   uzřěv,   ano pohani   veliký hod   činie
            see-CVB.PT ANO pagan-NOM.PL great   feast-ACC.SG bereiten-3PL
            '(as) he saw that the pagans prepared for a great feast.'

       b.   uzříte buoha, an divné   divi   činí
            see-2PL god-ACC -an miraculous miracles-ACC.PL tu-3SG
            'You see god, how he does wonderful miracles'

                                              (Pass., 2nd. h. 14th c., cit. Zubatý 1909)

This distribution would come out for free under a relativizer analysis (Gebauer 1903); but Zubatý (1909, 116) and Trávníček (1930, 116f) seek a different explanation: They assume that (10-b) was diachronically derived from a structure like (10-a) via an intermediate step of *uzříte buoha, ano divné divi činí* by a concomitant rise of agreement. Zubatý adds that the latter structure depended on a coreferent interpretation of *an/on* and was thus possible only after *on* had changed into a personal pronoun; Trávníček strongly disagrees with this detail, being convinced that Cz developed a weak 3P pronoun only later (see below section 4.3.1.2). There are some potentially problematic early and many later (18th century) examples of *an* (MASC.SG) occuring with (non-agreeing) overt embedded subjects (Trávníček 1930, 112), but they may partly be due to deletion of the final *-o*, as in *tak < tako, jak < jako, tam < tamo* (Trávníček 1961, 56). *Ano* is also used as a connector for embedded impersonal and indefinite-

personal clauses (Trávníček 1961, 51f).

(11)   a.   Že   jest   tak těžko pustiti od   světa,   ano jest   tak pracně
               that be-3SG so   hard   leave   from world-GEN.SG ANO be-3SG so   tiring
               slúžiti světu!
               serve   world-DAT.SG
               'That it is so hard to leave the world, yet so tiring to serve it!'

                                                (ŠtŘb, end 14th c., DČNK)

    b.   těch velblúdóv   na wssieczky strany   města, kamž se   obrátíš,
               those camel-GEN.PL on all   side-ACC.PL city-GEN where REFL turn-2SG
               ano sie   s   nimi   wssiady ženú
               ANO REFL with they-INS chase-3PL
               'those camels are in every corner of the city, wherever you turn, behold, ever-
               ywhere they chase around with them'   (Kabátník, 1491, DČNK)

Schematically, the theory advanced by Zubatý (1909) and extended by Trávníček (1930, 1961) may be summarized as follows, with *ano* standing for the non-agreeing form, and *an,-a,-o* for the agreeing one:

(12)   a.   *main clause:*
               1:    *particle + separate clause*   [$_S$ *ano*] [$_S$ ...]
               2:    *integrated particle*   [$_S$ *ano* ...]
               3:    *agreement*   [$_S$ *an,-a,-o* ...]
    b.   *embedded clause:*
               2:    *subjunction*   [$_S$ ... [$_{S=CP}$ *ano* ...]]
               (3:)   *embedded subject*   [$_S$ ... [$_{CP}$ *ano* ... NP$_i$ ...]]
               (4:)   *coreferential null subject*   [$_S$ ... NP$_i$ ... [$_{CP}$ *ano* ... pro$_i$ ...]]
               (5:)   *agreement*   [$_S$ ... NP$_i$ ... [$_{CP}$ *an,-a,-o* ... pro$_i$ ...]]
               (6:)   *relative clause*   [$_S$ ... NP$_i$ ... [$_{CP}$ Op$_i$ *an,-a,-o* ...]]

From Step 1 to Step 2, integration into the clause takes place, changing the category of *ano/ono* from interjection to particle and complementizer, respectively. This is the end of the derivation for adverbial clauses. Step 3 shows the beginning development towards a relative clause; step 4 the establishment of coreference; step 5 the rise of agreement; and step 6 the final stage of the relative clause.

It must be noted that the actual evidence for step 4 in (12) is remarkably sparse: Trávníček (1930, 112) mentions only three examples, one of which is

(13) slyšie ďábelské hlasy,   ano je   volají jich vlastními jmény
      hear-3PL diabolic voices-ACC ANO them-ACC call-3PL their own-INS.PL names-INS
      'They hear diabolic voices which call them by their own names'

(Mill., cit. Trávníček 1930)

Another one of the three contains non-agreeing *an* with a feminine NP$_i$, which *might* be due to deletion of final *-o* (and thus relevant) or just as well an early petrification of agreeing *an* (see below). Since the approaches discussed here all seem to care much more for a conceptual discussion of the changes than for their actual timecourse, as it is reflected in the data, none of them provides any quantitative diachronic evidence, and the fact that some of the hypothesized structures are almost non-existent, is willingly disregarded. In the case at hand, the idea of a gradual transition from non-agreeing *ano* towards agreeing *an* simply finds no empirical support. Relative clause structures like (10-b) seem to have contained agreeing relativizers exclusively since the beginning of written records.

On a more conceptual level, steps 4-6 of (12) contradict widespread views on grammaticalisation (cf. Lehmann 1995 and especially Diessel 1999): *Ano* allegedly changes from a fully grammaticalised, formally petrified complementiser to an agreeing, inflected relative pronoun and (according to Trávníček) further to a personal pronoun. This is the opposite direction of what is usually assumed for relativisers (cf. English *that*) as well as for pronominal and null subjects. Newmeyer (2001, 211) discusses Van Gelderen's (2000) analysis of expletive *there*, which developed from the Old English demonstrative *Þære* through relative pronoun to (non-agreeing!) expletive pronoun, and notes that "if *there* is a lexical noun, as is generally assumed, then we have an example of an upgrading from determiner to noun, that is, from a functional category to a lexical category".

4.3.1.2 *ono* **and** *on*

Zubatý and Trávníček do not systematically distinguish between the development of *ano* / *an* and *ono* / *on*. They assume that both went in a parallel or partially overlapping fashion, when *ano* was split into *a ono* etc. We will try and keep the two separate here, noting that there is a grey area in which it is impossible to decide which one is at stake.

Trávníček's criticism of Gebauer's view centers around the fact that both in his derivation of the relativiser and of the expletive, there is an intermediate step involving a coreferential pronoun. Actually, the nature of this pronoun is never discussed explicitly by Gebauer, and it remains somewhat unclear also in Trávníček's own work. Straka (1926/27; inaccessible to me, cited after Trávníček 1930, 122ff) wants to analyze *on*, in all its early uses (before the late 15th century), as an anaphoric *demonstrative*, similar to *ten*. He argues that on the

one hand, there was non-agreeing *ano* developing into a complementizer, and on the other, the agreeing relative pronoun *an, -a, -o*, consisting of *a +* an anaphoric demonstrative *on*, which later became separated again. So for this second development, the pronoun *on* was the driving force, and agreement would not need to arise, but would come "for free".

Trávníček (1930, 123f) tries to show that the anaphoric links assumed by Straka are also unwarranted in many examples, i.e., that there was no textual coreference involved at all. Note that this would be quite unusual for a demonstrative (cf. the case of *ten*). All authors except for Gebauer, but most explicitly Trávníček (1961, 29f), take the argument even further and claim that the modern personal pronoun *on* actually arose from the former *náladové 'emotive' on, -a, -o* by a loss of emotive overtones, where no agreeing associate was present. In this wording, the claim would imply that *ono* could lose "emotive overtones" also in impersonals; this is certainly unintended, as noted above in connection with (4-a). What Trávníček has in mind are cases of ambiguity between particle + referential pro-drop on the one hand, and coreferent subject pronoun on the other, which might then lead to the latter analysis. He demonstrates this on embedded cases like

(14) mňeli,   že spí,   a   on (Ø) byl   bez   duše
      thought-3PL that sleep-3SG and ON   was-M.3SG without soul-GEN
      'They thought that he slept, but (see,) he was dead.'

<div align="right">(FrantPr, cit. Trávníček 1961, 30)</div>

Note that under the "coreferent subject pronoun" analysis (with *on* being the subject and no null category present), *on* could not be classified as a weak pronoun, but only as a strong one, in standard Cz – modern Cz does not have weak pronouns (cf. also section 3.4 and Fehrmann & Junghanns 2008). Typologically, such a development from expletive to pronoun would certainly be unusual: French *il*, English *it* and *there* were pronouns before becoming expletive (Roberts 2007 and references therein; Gelderen 2000); cf. also Lehmann (1995) and Diessel (1999) for the claim of a grammaticalization path. Be that as it may, it is interesting that the idea actually presupposes a change *towards* a weak personal pronoun *on* in Cz. Given the clear canonical pro-drop properties of the modern language, this is already unexpected all by itself, and as we saw in section 3.4 (cf. also Fehrmann & Junghanns 2008), there are good reasons to assume that modern standard Cz has no weak pronouns. To add to the confusion, Trávníček (1930, 135) *does* assume a personal pronoun *on* even for the earliest periods, but only for cases of contrast with other pronouns (*ty* vs. *on*; *jiný nežli on* etc.) – i.e., a strong pronoun was already available anyway.

Despite artful rhetorics, the precise reason for agreement to become obligatory remains

somewhat unclear in Zubatý's and Trávníček's accounts. Zubatý (1909, 117) states in passing that even for *ano < an*, analogy with sentences containing the personal pronoun *on* may have played a role, which both Straka (1926/27) and Trávníček (1930, 128) contend – strangely enough, given that they otherwise deny the existence of such a pronoun in early Old Cz. Trávníček later hypothesizes that *ano, ono* as complementizers might have changed in analogy to agreeing relative pronouns like *jenž*, albeit only formally, not semantically (on his view, they were never pure relativizers, cf. also Trávníček 1961, 55). For the emotive *ono* in main clauses, such an explanation is unavailable anyway; here Trávníček (1930*,* 129) simply assumes that non-agreeing *ono "bylo jazyku nepohodlné, protože odporovalo běžným normám"* – *"was uncomfortable for the language, because it contradicted current norms"*; later, he mentions avoidance of similarity with the newly arisen answering particle *ano* as a motivation (1961, 27). The idea that agreement had to come about by the general laws of the language, is, of course, at issue with the expletives of English, French, and German, and with Czech *to*, which all lack gender and number agreement.

A second problem is that *ano* and *ono* indeed behaved differently with respect to agreement. While non-agreeing *ano* remained vital for centuries (see below), there is virtually no evidence of the transition from non-agreeing *ono* to agreeing *on* – exactly as above in connection with *an*-relatives. Zubatý himself found only three relevant examples, all in combination with the clitic modal particle *-(ž)ť* (15-a) and stemming from the translation of Aesop's fables (earliest preserved redaction 1472 in the *Sborník hraběte Baworowského*). Trávníček (1930, 60) adds one more from the St. Vít's manuscript (ca. 1380-1400) (15-b):

(15)   a.   Onožť   kuóň   jé,   travú   chodě, po   lúce   zde i   onde
                ONO-MP horse-NOM eat-3SG grass-INS go-CVB   over meadow-DAT here and there
                brodě
                walk-CVB
                'Behold, the horse is eating, walking through the grass and here and there over the meadow.'   (Ezop 2056 48a, cit. Zubatý 1909)

        b.   a   ono buóh   učinil   všecky věci   dobré   velmi
                and ONO god-NOM make-PT.M.SG all   thing-ACC.PL good-ACC.PL very
                'And behold, the Lord made all things very good.'
                                      (Vít. 96b, cit. Trávníček 1930)

To further support their claim, they cite dialectal evidence for non-agreeing emotive *ono* from Eastern Bohemian (16-a) (Zubatý 1909, 67), Moravian (16-b) (Trávníček 1930, 69) and Slovak (16-c) (late 19th/early 20th century) spoken examples:

(16) a. A   vono to   byl   velkej   kus   papíru
        and   ONO   that be-PT.M.SG big-M.SG piece-M.SG paper-GEN.SG
        'And that was a big piece of paper.'   (cit. Zubatý 1909, 67)

    b. ono pohanka   je pořezná   věc
        ONO buckwheat-NOM is cuttable-NOM thing-NOM
        'See, buckwheat porridge is easy to cut'   (cit. Trávníček 1930, 68)

    c. dnes bude   (muž)   celý deň fučať, ... že   ho   moria
        today AUX-FUT.3SG man-NOM all day grumble   that he-ACC vexate-3PL
        hladom. A ono nik   ho   nemorí.
        hunger-INS and ONO nobody he-ACC NEG-vexate-3SG
        'today my husband will grumble all day, ... that he is being vexated by hunger.
        But see, nobody is vexating him.'   (cit. Trávníček 1930, 68f)

Note that the Slk (16-c) is probably irrelevant here, because the linear order of the 2nd position clitic *ho 'him'* shows that *ono* is not integrated into its clause, but forms a separate unit, an interjection.

Finally, Zubatý (1909, 68-75) provides an interesting collection of related cases in other early Slavic languages: *Ano, ono* and *on* conforming to the Old Czech distributional rule are attested in the Old Polish *Biblia królowej Zofii* and in the sources of Linde's (1854-61) dictionary, and it is very unlikely that all of them can be explained away as Bohemisms. Old Polish also knows the demonstratives *owo* and *oto* along with *ano, awo* and *ato*, so the *a+o+demonstrative* morphology seems to be completely regular. Old Russian documents contain scarce examples of *анъ* and *ано*, which already Sreznevskij (1893/1955) relates to Old Czech *ano* (cit. after Zubatý 1909, 70ff). Interestingly, uses akin to the Cz expletive existed in Old Ru, and in impersonals even persisted up until the period of modern Russian – cf. (17-a) from our corpus, and (17-b) from recent work by Skorniakova (2006, cit. after Szucsich 2008):

(17) a. и они   Нѣмцы   учали   бить челомъ   Государю
        and ON-NOM.PL German-NOM.PL learn-LPT.PL beat forehead-INS Lord-DAT
        'and behold, the Germans learnt to bow before the ruler'

                                (Pskov chronicle, 14th c., cit. Zubatý 1909)

    b. Ono teplo-to   teplo,   da metet.
        ONO warm-ADV-MP warm-ADV but blow-3SG
        'It is warm, but it snows heavily.'   (modern Ru, cit. Szucsich 2008)

Despite the suggestive comparative and dialectal evidence, it still seems fair to say that the empirical case for a diachronic development of *ono* towards agreeing *on* in Old Cz is not very strong. We will return to the Old Pol case below.

4.3.2 Old Cz corpus evidence and its implications

As noted above, despite the ingeniousness and empirical richness of the studies by Zubatý (1909), Trávníček (1930; 1961) (Z&T in the remainder) and others, neither the time course of the alleged changes nor the quantitative amount of evidence have been sufficiently taken into account, so that the theories remain somewhat speculative. Furthermore, the case of closely related languages is mentioned, but not scrutinised in any detail. Therefore, it does make sense to carefully reconsider corpus data and weigh the arguments for and against the various approaches. Possibly, some of the conceptual and argumentative "blind spots" criticised above can be clarified along the way.

4.3.2.1 "Emotive" *ano, ono* and the rise of agreement

One of the central differences between Z&T and the analysis by Gebauer is that the latter favours a scenario in which an agreeing pronoun develops into a relativiser and further into a general, uninflected subordinator, while the former envisage a change *towards* agreement by which the former interjection becomes a relativiser (subordinate clauses) or an emotive expletive (main clauses). Unfortunately, evidence for the "rise of agreement" is insufficient. One would expect to find a clear difference between early texts which contain mostly *ano, ono* and later ones in which these elements changed into *an, on.* In fact, the earliest examples of emotive *ano* date from the second half of the 14th century (Trávníček 1930, 66, citing Straka). Among these are already instances of reflexive clitics adjacent to the particle:

(18)   ano sě   jim   těžko   k súdu   ukázati
       ANO REFL they-DAT difficult to court-DAT show
       'Behold, it is difficult for them to show up in court'   (ŠtŘb, 1392, DČNK)

Although *sě* had certainly not become restricted to 2nd position yet (the fixation of its position took place in 19th-20th c.), it was nevertheless a true clitic and could not occur after a pause or sentence boundary. Thus, *ano* in (18) serves as a clitic host and is therefore clearly syntactically integrated. However, we cannot exclude that we are really dealing with agreeing *an* in an impersonal environment here; agreeing emotive *an, -a, -o* was already available around the same time (contrary to Z&T):

(19)   [Takého-li tobě slušie v zalarz držěti?]

       An   by   sě   viece na vašeho   buoha   rozhněval   a   vešken by
       AN   COND REFL more   on your-ACC god-ACC be-angry-LPT.M.SG and whole   COND
       jej   rozbil.
       it-ACC destroy-LPT.M.SG

'[Is it decent for you to keep such a person in the carcer?] That one would be even more angry about your God and would destroy it (the carcer) completely.'

<div align="right">(Pass., 2nd h. 14th c., DČNK)</div>

Thus, if there was a phase of *exclusively* non-agreeing *ano*, it must have been before written records. In the available documents, we simultaneously find non-agreeing, syntactically integrated *ano* and agreeing *an*, but no clear evidence of a development from one to the other. Non-agreeing *ano* still remains perfectly common for several centuries, both as a subjunction and as a main clause particle (the latter developing semantically), cf. (20):

(20)　a.　a　vy,　křesťané,　…　nepatříte,　ano　bratry　vaše
　　　　　and you　Christians-voc.pl　　neg-observe-3pl　ano　brothers-acc.pl　your-acc.pl
　　　　　zřejmě　jímají
　　　　　obviously　catch-3pl
　　　　　'And you, Christians, … do not even see how they catch your brothers'

<div align="right">(BartHla, DČNK, 1595)</div>

　　　b.　Ano　hynou　mnohé　byliny, jestli je　　vodou,　v　kteréž　…
　　　　　ano　die-3pl　many　herbs　if　they-acc.pl water-ins　in which-loc
　　　　　pokropíš
　　　　　spray-2sg
　　　　　'Behold, many herbs die if you spray them with water in which …'

<div align="right">(Hosp, DČNK, 1705)</div>

The situation with *ono, on* is even worse for the Z&T account: Emotive *ono* is attested no earlier than the end of the 14th century (1380-1400), cf. (15-b). As mentioned above, this case actually represents an absolute rarity in Old Cz – it does not occur in the extended corpus data presently available (which includes both the DČNK and the extended corpus of the *Vokabulář webový*) at all. Trávníček (1961, 25) states that emotive agreeing *on* – like *an* – became "more frequent" only from the 15th century onwards; but he himself presents several examples from the *Hradecký rukopis* (around 1360); a similar case is

(21)　On jistý　　　Krok　zpravoval jest zemi
　　　on certain-nom K.-nom rule-lpt.m.sg aux earth
　　　'Behold, a certain Krok ruled the world.'　　　　　　(PulkKr, ca. 1400, VokWeb)

The alleged change from *ono* to *on*, which would have had to take place during the 15th century, loses plausibility when the relatively numerous *earlier* instances of agreeing *on* are taken into account. We cannot disprove that there was a phase without agreement before written records; and it is just as well possible that agreement was optional first and became

obligatory later. But the change would have had to be so profound that no trace of non-agreeing "emotive" *ono* survived. In some sense, this is the reverse problem from the one with *ano*; what one would like to see as evidence of a change are *some* remains of the original pattern, rather than all (*ano*) or none (*ono*).

According to Z&T, agreement came about as a direct consequence of the *integration* of *ano, ono* into the following clause. However, integration is clearly no sufficient condition, since there are systematic cases of integration without agreement in Old Cz (22-a), and diachronically, agreement is usually lost rather than gained (cf. the earlier manuscript of the same text in (22-b)):

(22) a. Vršovici, ješto   byli kněží svadili, těch   sú Moravěné
          V.-NOM.PL that-NEUT.SG AUX Lord's disunited those-GEN.PL AUX M.
          mnoho zbili
          many   beat-PST
          'The Vršovici, who had disunited the Lord's (people), of those the Moravians beat up many.'

                                                        (DalL, 15th c., cit. Bauer 1967)

      b. Vršovici, již   ...
          V.-NOM.PL who-NOM.PL

                                                        (DalC, mid 14th c., cit. Bauer 1967)

Moreover, it is difficult to make sure that *ano, ono* was in fact syntactically integrated into its clause, i.e., that there was no pause (clause boundary) between the two – even more so in written, historical texts. Lindseth (1998) acknowledges this problem, but decides to follow Trávníček's authority nevertheless. Quite a few potentially relevant examples can be found in the modern Sl[ova]k National Corpus (but next to none in the Czech National Corpus), cf.

(23) Ono ten dobový účes   robí   svoje.   [Slk]
      ONO this modern haircut-NOM.M.SG do-3SG POSS.REFL-ACC.N.SG
      'This modern haircut does it.'   (SNK)

The three Slovak native speakers whom I consulted about such examples, found them more or less acceptable only with a slight pause after *ono*, i.e., without clausal integration. It is obvious that the evidence for the "rise of agreement" in the case of *ono, on* is not only sparse, but also very fragile.

4.3.2.2 *ano, an* **as complementiser and relativiser**

Since the earliest written records, *ano* is the preferred complementiser for verbs of perception such as *viděti 'to see', uzřieti 'to make out'*, but also e.g. *nalézti 'to find'*, with non-agreeing and agreeing forms occuring around the same time and often within the same text:

(24) a. svýma   očima   viděla,   ano tvój milý syn   velikým
          REFL-POSS.INS.DU eye-INS.DU see-LPT.F.SG ANO your dear son-NOM great
          hlasem   zavolav
          voice-INS.SG call-CVB.PT.ACT
          'with her own eyes, she saw how your dear son shouted in a loud voice'

     b. když-s jej   na kříži   viděla,   an krví   brocen a
          when-2SG he-ACC on cross-LOC.SG see-LPT.F.SG AN blood-INS.SG redded and
          má   rány   v rukú
          have-3SG wound-ACC.PL in hand-LOC.DU
          'when you saw him at the cross, as he was red of blood and had wounds in his
          hands'   (Milič., DČNK, 2nd h. 14th c.)

As noted above, *ano* and also *an* remain vital in this function for a long time; in 17th c., *ano* gets used mainly as a particle, e.g. as a focaliser (25), since the 19th c. also as a modal and answering particle.

(25) Poněvadž Pán Krystus netoliko   jest   učiněn   dokonalým mužem,
      namely   lord K.   not-so-much be-3SG make-CVB.PT.PAS perfect-INS man-INS.SG
      ... ano i   maličkým děťátkem
          ANO also small-INS.SG child-INS.SG
      'Christ our Lord has not only been made a perfect man, [...] but also a little child.'
                                                                    (HavManž, DČNK, 1615)

Since the early 17th c., we often find *an* without agreement as an adverbial clause subordinator, before the agreeing variant is 'revived' in 19th c. (Jirásek). With these and other verbs, the agreeing variant can often be paraphrased as a relative clause, but this is not always the case:

(26) Tehda opět hlas   dietěcí   vzvola,   an vyšed,   opět
      then   again voice-NOM.SG children's-NOM.SG call-AOR.3SG AN go-out-CVB.PT.ACT again
      nikoho   nenalezl.
      nobody-ACC find-LPT.M.SG
      'Then again a children's voice called, and he (=Kristoforus), going outside, again did
      not find anybody.'   (Pass., DČNK, 2nd h. 14th c.)

In (26) and several similar cases (cf. also (14) above), *an, -a, -o* indeed behaves like a combination of a subordinator and a *strong* pronoun *on.* This fact is problematic both for Z&T's derivation steps 4-5 in (12), but also for Gebauer's (1970; 1929/2007) view: The embedded subject in (26) could neither be *pro* nor an overt weak pronoun. Straka's (1926/27) analysis (cited after Trávníček 1930, 122ff) as *a + demonstrative pronoun* seems most adequate here. But what was the actual status of the 3P pronoun in Old Cz? In section 3.5.2.2, we saw that there is no evidence for any radical change between Old Cz and Modern Cz apart from minor adjustments: Both behave like canonical null subject languages, with the 3P subject pronoun being strong. If any, we could identify a few weak uses in the Old Cz corpus. Specifically, no diachronic development towards a weak personal pronoun could be detected (contra Trávníček 1961). Following Diessel (1999, 99), Givón (1983), Himmelmann (1997) and others, the specific typological difference between demonstratives and personal pronouns in context consists in the fact that the former, but not the latter, induce a shift of topic. Kresin (1994, 98) and Uhlířová (1987, 29) state that this is exactly the case with *on* in *modern* (standard) Cz: *On* cannot just be coreferential with preceding subject or topic, but always induces a contrastive or deictic interpretation. I conclude that the idea that modern Cz *on* developed from the expletive via a loss of emotive features (Trávníček 1961, 29f) misses the point.

4.3.2.3 The an(o) > a+on(o) split

As before, diachronic evidence for such a process is simply missing: As (21) shows, *on* as agreeing, clause-internal expletive was already available in the 2nd h. 14th c., i.e., at the same time as "emotive" *ano* and *an. On, an* and *a on* could easily co-occur in the same text:

(27) A   on   jej   druhé naplni,   an mu   sě   opět převráti.
      and he-NOM it-ACC second fill-AOR.3SG AN he-DAT REFL.ACC again turn.over-AOR.3SG
      A   když jej   po třetie naplni,   on mu   sě   opět převráti.
      and when it-ACC at third fill-AOR.3SG he he-DAT REFL.ACC again turn.over-AOR.3SG
      'And he filled it (= the mug) a second time, and he turned over again. And when he
      filled it a third time, it turned over again.'   (ŽSvO, 1st h. 15th c., DČNK)

Conceptually, it is also strange that *a + on(o)* should first merge into *an(o)* before the written period, in order to get separated again later. If there was a way of deriving the properties of *ono* without recourse to *ano,* this would clearly be preferable. Before I propose such an alternative, let us shortly look into the closely related case of *ano, ono* and *on* in Old and Middle Polish.

4.3.3 Old and Middle Polish evidence

As mentioned above, Zubatý (1909) observes several parallel cases of *ano* and/or *ono* in similar functions as in Cz, in other early Slavonic languages. We will focus on Pol here, which, of course, does not have an expletive *on* any more today, but did so in history.

4.3.3.1 *ano, an*

As noted in several historical grammars and dictionaries, *ano* and *an, -a, -o*, contracted from *a+on(o)*, occurred as sentence connectors in Old Pol roughly until the turn of the 15th-16th c. (Klemensiewicz 1999, 122; Brückner 1998; Pisarkowa 1984, 230). Pisarkowa distinguishes the functions of (i) a paratactic connector, (ii) an adversative subordinator, and (iii) a complementiser for verbs of perception. The latter may be extended to a general *formuła obecności 'formula of presence'* which mainly points to the simultaneousness of main and embedded clause actions:

(28)   Gdyby ujrzał   psa,   a   on za   tobą idzie
           if   see-LPT.M.SG dog-ACC.SG and he behind you   go-3SG
           'If he makes out a dog as it is going behind you [...]'

<div align="right">(BielŻywFil, 1535, cit. Pisarkowa 1984)</div>

The *Słownik staropolski* (Urbańczyk 1953-2002) lists both non-agreeing *ano* and agreeing *an* in the same three functions as Pisarkowa (1984) above, i.e., as paratactic connector synonymous to *'a on, i oto'*, as adverbial subordinator meaning *'ale on' (an)* or *'a jednak, a tymczasem, bo właśnie' (ano)*, or as temporal simultaneousness marker *(an)* or complementiser *'że właśnie, jak właśnie'* (*ano*):

(29)   Owa gdyż wynidę   na pole,   ano zbici   mieczem, a
           PTCL when go-out-1SG on field-ACC ANO killed-CVB.PT.PAS.NOM.PL sword-INS   and
           jestli wnidę   do miasta, a   onić   zginęli
           when go-in-1SG in   city   also these-NOM.PL perish-LPT.PL
           'If I go forth into the country, behold, those slain with the sword, or if I enter the
           city, behold, those also perished.'   (Jer.14:18, BZ 1455, PolDi)

The agreeing version *an, -a, -o* must have been rare in Old Pol, as there is not a single example in the *Korpus staropolski*, despite numerous instances of *ano*. Accordingly, the *Słownik 16. wieku* (Pepłowski 1966-2004) lists rich evidence for *ano*, but does not even have an entry for *an* any more. As in Cz, *ano* came to be used as a confirming or demonstrative particle:

(30)   iákoby zwodnicy   áno prawdźiwi: iáko nieznáiomi áno
          as-if   impostor-NOM.PL ANO truthful   as   unknown   ANO

      znáiomi: iáko vmieráiący a   oto żywiemy
          acquaintance-NOM.PL wie   dying-NOM.PL and OTO live-1PL

      'as impostors and yet (are) true; as unknown, and yet well known; as dying and

      behold, we live'   (2nd Cor. 6:8-10, Wujek 1593, Sł16)

Incidentally, the *Biblia Brzeska* (1563) uses *ale* instead of *áno* here; the *Biblia Gdańska* (1632) has *wszakże*; modern bible translations use *jednak* (Biblia Warszawska 1975. Corpus examples from PolDi (16th c.) demonstrate that non-agreeing *ano* was clearly treated as clause-internal, because it supports clitic pronouns:

(31)   jako widział   przez sen,   *ano mu*   król   niemałą sumę
          as   see-LPT.M.SG through dream ANO he-DAT king-NOM not-small sum-ACC.F.SG

      czerwonych złotych   daje.
      red   Złoty-GEN.PL give-3SG

      'how he dreamt that the king gave him a remarkable sum of red Złoty'

                                          (Dworzanin, 1566, PolDi)

The *Słownik polszczyzny 17.-1. poł. 18. wieku* (Siekierska 1999-2004) lists neither *an* nor *ano* any more. But according to the *Słownik gwar polskich* (Karas et al. 1977-2005), *ano* was still present in 20th c. dialects (without a clear regional delimitation) as a modal particle, a deictic interjection or paratactic, adversative complementiser.

4.3.3.2 *ono, on*

Pisarkowa (1984, 29) found that *ono* could occur as an expletive in 16th c. Pol, albeit rarely: *"Rzadkim archaizmem, który utrzymał się w niektórych gwarach, a w polszczyźnie literackiej poświadczonym tylko w pierwszej połowie XVI w., jest zaimek* ono *jako podmiot formalny, treściowo nieokreślony:"* – *'A rare arcaism, which survived in some dialects and is attested only in the first half of the 16th century in literary Polish, is the pronoun* ono *as a formal, semantically undefined subject'.* An example from PolDi is provided below:

(32)   Także też ono jeden pirata, [...] przywiedzion   do Aleksandra
          so-that also ONO one   pirate   lead-CVB.PT.PAS.M.SG to   A.-GEN

      Wielkiego
      Great-GEN.SG

      'So that also a pirate [...] was brought before Alexander the Great.'

                                          (Rej: Żywot, 1567, PolDi)

The evidence provided by historical dictionaries is actually not so sparse: The *Słownik staropolski* (Urbańczyk 1953-2002) mentions numerous examples of *ono* in the meaning of *oto (tam)* or in an emotive use, overlapping with items like *owa, ować, owo, owoć, onoć*. The *Słownik 16. wieku* (Pepłowski 1966-2004) describes *ono* (i) as a deictic particle referring to a proposition as a whole or to one of its parts, and (ii) as a subordinator following perception verbs (359 instances altogether). The two subtypes, of course, correspond exactly to the expletive vs. subordinating subtypes of *ano, an* and *ono, on* in Cz. Attributive demonstrative *on*, apart from coreferential and deictic readings subclassified in many ways, also has a purely "emphatic" readings (3818 examples) which may highlight not only its nominal associate, but also larger phrases. Examples are very close in meaning to the emphatic expletive (ex. after Pepłowski 1966-2004, vol. 21, 430ff and 441ff):

(33)   Ono   wykładacż   snow   przyszedł.
       behold dream-reader-NOM again come-PT.M.SG
       'Behold, the dream reader is coming again.'   (Gen. 37:19, BielKron, 1564, Sł16)

(34)   A   *oná* śmierć   iáko chłop dybie s   tyłu   s   kosą.
       and ONA death-NOM.F.SG as farmer cut-3SG from behind with scythe
       'And behold, death cuts from behind like a farmer with a scythe.'

                                                          (Rej: Zwierzyniec, 1562, Sł16)

The fact that *ono* need not agree is remarkable from the perspective of Cz. Furthermore, agreeing *on* as well as non-agreeing *ono* could occur in a deeper clause-internal position:

(35)   A   dźieie   sie częstokroć v   nas   z   cżego   sie *ono* pośmiewał
       and happen-3SG REFL often   in we-PRP from what-GEN REFL ONO laugh-LPT.M.SG
       Diogenes
       D.-NOM
       'And often among us happens that about which Diogenes laughed.'

                                                          (PowodPr 1595, Sł16)

These examples show that there was an expletive-like, "emotive" *ono* and also a purely emphatic attributive demonstrative *on* present in Pol at least throughout the 16th c. What apparently was not found is a *separate* emphatic *on*, discontinuous with its associate.

The characteristics of the personal pronoun *on* were shown above (section 3.5.2.3.2) to have been rather stable over time. Differences in use in the parallel bible corpus concern changes in the meaning and use of single demonstratives and in some instances of topic shift, which, however, do not point uniquely in the direction of introduction or loss of weak pronouns. What was clearly lost was attributive *on* (= *ów* in modern Pol).

To summarise, Old Pol, up to and including the 16th c., used non-agreeing *ano* as a complementizer in adverbial clauses and perception verb complements, and as a confirming or demonstrative particle. 16th c. Pol also knew non-agreeing and agreeing *(a) on(o)* as a demonstrative particle, which got integrated into the clause which it modified. *Ano* was lost before the 17th c., as well as probably *ono, on* in the relevant meaning. In the present context, the crucial item missing is agreeing *an* – a fact confirmed by corpus research based on the PolDi corpus; instead, there is quite some evidence of non-agreeing *ono*. This presents a problem for approaches which rely on the relativizer as an intermediate step (Gebauer and Straka) as well as for those which generally count on a "rise of agreement" (Z&T): If *ano* and *ono* in Old Pol never began to agree, then agreeing *on* cannot stem from this source.

4.3.4 The rise of expletives in Czech and Polish: An alternative account

We have seen that the Z&T proposal lacks empirical support in the following respects: (i) There is next to no evidence for non-agreeing *ano* + lexical subject or *ono* + lexical subject. Therefore, a diachronic "rise of agreement" cannot be proven. (ii) *An* in embedded clauses cannot always be a relativizer, but sometimes clearly involves *a* + strong pronoun. (iii) The development envisaged by Z&T does not reflect the distribution of emotive emphasis according to Lindseth (1998) and Hirschová (1984): Agreeing *on* should be colloquial, but *not* emotive/emphatic, while only non-agreeing *ono*, as it occurs in (truly) impersonal structures with expletive *pro*, should be emotive/emphatic. (iv) Evidence for the split of *ano* into *a ono* and of *an* into *a on* is missing. (v) Old Pol, which is cited by Z&T as a parallel case, interestingly, had only *ano, ono*, and *on*, but it never developed agreeing *an* or banned non-agreeing *ono*. There was thus no "rise of agreement", but crucially, expletive-like uses of agreeing *on* did exist, with solid frequency, throughout the 16th c. They must have a had a different source.

Of course, there is no way to prove that the alleged changes cannot have happened before the earliest written documents, and thus, Z&T's theory cannot be refuted – apart from the clearly erroneous claim that a weak pronoun on developed from the expletive later. One might also assume that the Old Pol case is simply independent and irrelevant for Old Cz. Nevertheless, I would like to propose an alternative scenario which (i) requires no empirically unsupported steps, (ii) fares better with the observations by Lindseth (1998) and Hirschová (1984), and (iii) allows for a unified view on Old Cz and Old Pol.

The crucial idea is to give up on a *unique* source of *ano, an, ono*, and *on* – as, in fact,

already intended by Straka (cited after Trávníček 1961). Instead, I assume that there was (i) the demonstrative particle *ano*, which developed into a complementiser and gave rise to non-agreeing *a ono, ono*; and (ii) the *attributive* demonstrative *on*, which could get separated from its head noun by a perfectly common movement process, be re-analysed as an expletive DP in the subject position SpecT, and optionally merge with the complementiser *a* into some sort of relativiser, not identical in function to *bona fide* (weak) relativisers.

Formally, the integration of *ono* into the following clause – as proposed by Z&T – would involve two restructuring operations (Meyer 2009a): First, into the C projection and then further "down the tree" into the T projection. Note that grammaticalisation in the sense implemented formally by Roberts (2007a) does not foresee such shifts "down the tree", but only "upwards", from lexical to functional head positions. This corresponds to the principles of Grammaticaisation theory.

This idea has the virtue of explaining the apparent subject "doubling", the agreement, and restriction of "emotional" emphasis to non-agreeing (impersonal) expletives. Regarding the latter issue, note that such a derivation is unavailable if there is no overt associate in the first place; then the only remaining analysis for *ano* (*ono*) is that of a demonstrative particle.

There is good evidence in favor of a correlation between the rise of expletive *on* and the existence of attributive *on*. Note first that the M.NOM.SG is the only distinct form in the two paradigms. While the modern Cz attributive demonstrative is unambiguously *onen* in M.NOM.SG, we find *on* as an attributive demonstrative both in Old Pol and in Old Cz (cf. also Gebauer 2007, 242). Pol had no *onen* and lost attributive *on* roughly during the 16th c.

(36) widźi on Bog [= deus ille] / złośćiwey tey ćierpliwośći wafzey
      see-3SG ON God   evil   that patience   your-GEN.SG
      złośćiwe przyczyny
      evil   reason-ACC.PL
      'That God sees the evil reasons of your evil patience.'   (ModrzBaz, 1577, Sł16)

In the Old Cz corpus data, *onen*, in fact, exceeds attributive *on* in frequency already in the 16th c. But since expletive *on(o)* is already well attested at that time, this is actually not an obstacle to the present proposal. The remaining instances of attributive *on* are indicative of an earlier form. Examples include

(37) a. A   král   ... sěka   on   ľud   hlúpý
          and   king-NOM   cut-AOR.3SG ON   people stupid
          'And the king ... cut that stupid people.'   (AlxB, ca. 1350, VokWeb)
      b. ženúc   ot   krále   onoho   /... s   člověkem ... jehož   on
          chasing from king-GEN that-GEN.M.SG   with person-INS   who-ACC   ON

> král   vypovědě   (AlxB, ca. 1350, VokWeb)
> king-NOM expel-AOR.3SG
> 'chasing away from that king … with a person … whom that king had expelled'

c. kdež sě on ľud kryjě lazí
   where REFL ON people hide-CVB creep-3SG
   'where this people, hiding, creeps'   (AlxM, ca. 1350, VokWeb)

The example in (37-a) clearly contains attributive *on*; in (37-b,c), however, an expletive analysis would not be absolutely excluded. In fact, about two thirds of my examples of expletive *on(o)* from 14th-16th c. Old Cz could also be analyzed instead as involving an attributive demonstrative adjacent to the subject DP. In the 16th c., however, only about one third of the examples would support such an analysis.

The most obvious apparent counterargument to the present proposal, possibly also a major reason for its not having been considered yet, comes from semantics: Expletive *on* clearly differs in meaning from demonstrative, attributive *onen*. The latter simply points to an entity remote from the speaker. The former, on Trávníček's view, comes with a special emphasis and emotive value. But note that the assumption of a semantic change from a demonstrative *pronoun* to a more or less semantically empty expletive is by no means unusual, cf. expletives like English *there* or German *da*. The semantic distinction drawn by Lindseth (1998) and Hirschová (1994), in fact, falls out nicely from the present proposal: Only agreeing *on*, being historically based on an attributive demonstrative pronoun, becomes virtually meaningless. The emphatic/emotional value of the non-agreeing original particle, however, would be retained.

The Old Pol data given above also fit in well: Attributive demonstrative *on* could give rise to the agreeing expletive, while *ano* and *a ono, ono* could still be used as particles. No general "rise of agreement", which would have excluded expletive-like *ono*, needs to be invoked. A true problem for the present proposal, however, is the lack such *non*-agreeing *ono* in Cz. I can only tentatively conclude that the reanalysis from attributive demonstrative to expletive, took place faster in Cz than in Pol.

At the theoretical level, the present proposal is somewhat at odds with Rezac's (2004) claim that the expletive and the associate *cannot* form a common constituent in modern Cz. As discussed above, however, this is not a necessary assumption anyway, because its effects are fully derivable from Binding Theory. A conceptual advantage may be seen in the fact that our proposal fits better with the changes known from grammaticalisation phenomena: Z&T imply a reversal of the earlier morphological fusion of *a+on(o)*, an increase of inflectional features and syntactic combinability, and a rise of lexical from functional categories. By contrast, the present proposal goes "up the tree", reducing positional and inflectional variability.

5 Indefinite Null Subjects

5.1 Introduction

Subjectless sentences with an indefinite interpretation are widespread among the world's languages (cf. Siewierska 2008a,b for a cross-linguistic view). Within Slavonic linguistics, traditional descriptive grammars have given the respective examples macro-labels referring to their configuration as a whole (like *neopredelenno-ličnaja forma, věty se všeobecným konatelem* etc.), but already Mel'čuk (1974) proposed a rich inventory of *null subjects* for Ru indefinite impersonal constructions. In Pol and Cz, indefinite impersonals are standardly marked by shifts in voice (reflexivisation). At least in Pol, null subjects can be shown – by their behavior-and-control properties, in Keenan's (1976) or Givón's (2001) terms – to be at stake in these constructions as well. As before, our main question is how the different expressions of indefinite impersonals came about.

The chapter begins with a very basic overview of the empirical landscape of indefinite impersonals in modern Ru, Pol and Cz, and a delimitation of the subclass which will concern us in the remainder. Next I turn to the current general linguistic view on these constructions, first in semantics-centered, then in syntax-centered approaches. An interesting cross-linguistic correlation between the availability of definite and indefinite null subjects, which has been recently advocated in the *pro*-drop literature, leads us to the issue of reflexivisation as a Voice phenomenon, and back to the analysis of reflexivisation and indefinite impersonals cross-Slavonically. In the central part of the chapter, I characterise the diachronic development of indefinite impersonals in Ru, Pol and Cz, applying the syntactic and semantic distinctions developed in the synchronic part.

Since indefinite impersonals can often be recognised only by semantic/contextual analysis, the corpus-based approach has to be partly indirect: On the one hand, I try and track down developments using the few overt exponents, namely, the reflexive marker and the 3sg/pl overt pronouns and verbal endings. On the other, large samples of partially pre-processed source text are evaluated, which I had previously annotated for the relevant categories. The results show (i) some clear developmental stages, and (ii) some rather recent developments which deserve to be scrutinised further in shorter-term diachronic studies.

5.2 Indefinite impersonals in modern Ru, Pol and Cz

5.2.1 Overview

It has been noted long ago that Slavonic languages use various constructions to express a general, "anonymous", inexplicit highest argument, a rough equivalent of overt markers like

German *man*, English *one*, or French *on* (cf. Kosta 1990, ch. 1.1.2 for the history of research on the topic). Semantically, these arguments are mostly restricted to human beings. Since they can get indefinite, arbitrary, or generic readings (for details see below), the label *"indefinite impersonals"* is somewhat misleading. It should therefore be taken as a convenient cover term – the opposite of "definite", so to speak, rather than truly "indefinite" in a semantic sense. The relevant configurations in Ru, Pol and Cz are

- 3PL, earlier or in varieties also 3SG, null pronominal subjects (1)

- reflexive verb forms, agreeing with an overt subject or non-agreeing in N.3SG, (2)-(3)

- impersonal passives, including the so-called *-no /-to*-construction in Pol (4)

- sentences with overt arbitrary subjects, mainly *člověk/człowiek/čelovek 'man'* or *jeden/odin 'one'*, cf. the example from Old Cz in (5)

(1)   a.   Zdes' ne kurjat.   [Ru]
          here   not smoke-3PL
          'One does/may not smoke here.'

     b.   Mówią, że cebula   ma siedem łupin. [Pol]
          say-3PL   that onion-NOM has seven   skins
          'They (gen.) say that an onion has seven skins.'   (Rytel-Kuc 1990, 138-145)

     c.   Sanitka   přijela,   vynesli   mamku.   [Cz]
          ambulance-NOM arrive-PT.F.SG carry.out-PT.PL mother-ACC
          'The ambulance arrived, mom was carried out.'   (Rytel-Kuc 1990, 138-145)

(2)   a.   Tu się pije   wódkę.   [Pol]
          here REFL drink-3SG vodka-ACC.SG
          'One drinks vodka here.'   (Rothstein, 1993)

     b.   *Tady se pije   vodku.   [Cz]
          here   REFL drink-3SG vodka-ACC

     c.   Tancovalo se až do rána.   [Cz]
          dance-PT.N.SG REFL until to   morning
          'There was dancing until the morning.'   (Růžička 1986, 249)

(3)   a.   ... způsoby, jakými   se   stížnosti   vyřizují,   ... [Cz]
              ways   which-INS REFL complaints-NOM manage-3PL
          '... the ways in which complaints are managed'   (ČNK)

     b.   Ta powieść   przyjemnie się czyta.   [Pol]
          that story-NOM.SG nicely   REFL reads-3SG
          'That story reads nicely/easily.'   (Szymańska 2000, 29)

     c.   *Piłka   rzuciła   się do wody. [Pol]
          ball-NOM.SG throw-PT.F.SG REFL in   water

                                (Szymańska 2000, 140)

(4) a. W budżecie przewidziano na ten cel 250 tys. zł. [Pol]
         in budget   reserved-NTF   for this goal
         '250000 zł. have been reserved in the budget to this end.'   (NKJP)

     b. Dostawano różne kary   i   bywano bitymi.   [Pol]
         received-NTF various punishments and be-NTF   beaten-INS.PL
         'One received various punishments and was being beaten.'   (Kibort 2006)

(5) když člověk netbá   na čest tohoto světa   [Old Cz]
     when man   NEG-care-3SG on honor this-GEN world-GEN
     'If one does not care for the honour of this world'   (Alx, 14th c., cit. Trávníček 1961)

These constructions are not distributed evenly across the respective languages: 3PL null subjects are the standard means to mark indefinite impersonals in Ru (1), but they occur a lot more rarely and with different semantic restrictions in Pol and Cz (Rytel-Kuc 1990, Berger 1991). In Cz, the reflexivisation "strategy" is predominant, while Pol uses reflexivisation or *-no*/*-to*-constructions. Relative frequencies of the varios indefinite impersonals differ: Rytel-Kuc (1990) compares the translations of 2917 German passives and *man*-clauses, and 131 Pol and 316 Cz sentences with a 3PL null subject, and 461 Pol *-no*/*to*-constructions. She also finds 6 Pol and 36 Cz sentences with 2P null subjects, an option well-known also from Ru (6). Some examples from older stages of Cz are given by Trávníček (1961, 34ff), e.g. (7):

(6) tiše   edeš'   – dal'še budeš' [Ru]
     quiet-CMP drive-2SG   far-CMP be-2SG
     'Most power lies in silence.'

(7) jak se   ho   dotkneš, tak se   zbodeš [Cz]
     as   REFL he-ACC touch-2SG so   REFL stinch-2SG
     'If one/you touch(es) it, one/you will stinch oneself/yourself.'

                                        (Jan z Lobkovic, 1515; cit. from Trávníček 1961, 34ff)

I will confine myself to 3SG/PL null subjects and reflexive constructions in this chapter. Previous work on the *-no*/*-to*-construction abounds, and, more importantly, borders on the issue of resultatives and periphrastic passives, which truly deserves a separate treatment of its own (cf. Wiemer and Giger 2005 for an overview, (Wiemer, to appear) for a recent account of Pol). I will also exclude from consideration the so-called *generalized-personal* constructions with 1PL or 2SG null subjects, cf. (6)-(7). This is done basically for reasons of space and coherence. The relation betwen these latter null subjects and the pro-drop properties of a given language is not so clearcut: Ru, which we classified as a partial pro-drop language, and Cz, a canonical null subject language, *both* have generalized-personal null subjects. This is no exception: Italian and Hungarian, both canonical null subject languages,

allow for generic readings of 2sG (Sigurðsson and Egerland 2009, 179). However, in another canonical pro-drop language, Spanish, the use of a null or overt pronominal subject in these cases strongly depends on dialect (Mayol to appear). By contrast, there are by now rather clear and theoretically modelled connections between 3sG/PL indefinite impersonals and pro-drop status.

Indefinite impersonals, as the term is used here, must be distinguished sharply from *true* impersonals. There are semantic as well as syntactic differences: While indefinite impersonals imply a *human* agent or experiencer per default, true impersonals carry no such restriction. One subclass may be understood as containing a natural force (Mel'čuk's 1974 *stixii*) or a semantically maximally bleached entity (Szucsich 2008) as causer, which does show syntactic effects like control into gerundial constructions. For other subclasses (e.g. meteorological verbs), traditional grammars assume no subject at all, while some formal approaches assume a null element, expletive *pro*, with no associated syntactic effects.

5.2.2 Semantic properties

The fine-grained semantic and pragmatic differences between the various expressions of indefinite impersonals have not received due attention in Slavonic linguistics yet. Berger (1991) reviews a number of accounts in the literature, and analyses the use of 3PL indefinite impersonals in Ru vs. Cz., with some remarks on Pol. He departs from a statement by Grepl (1973, 147), who states (i) that Reflexive and Periphrastic Passive in Cz exhaust the realm of semantic deagentivity, so that a 3PL null subject is unnecessary and occurs only colloquially, and (ii) that this difference between Ru and Cz is related to the difference in referential null subjects. Berger shows that the former is wrong: There are semantic specifics of the 3PL indefinite impersonals in Cz which make them non-synonymous with Reflexive or Periphrastic Passive. In our context, the second part of Grepl's statement is also interesting, since it reflects what one might plausibly assume as a first idea about canonical null subject languages: Since any other null subject could be mixed up with a referential one, it should be impossible. We will see in section 5.3.2 below that this is also generally wrong.

According to earlier descriptions by Panevová (1973), the 3PL implies that the Agent is a person and part of a collective which includes neither the speaker nor the hearer, so that a generalising (= generic) reading is impossible. Berger further cites Dušková (1973) with the observation that the 3PL null subject in Cz refers to a group of people specified by temporal or local delimitation, in the vast majority of cases. The same reading can be found in Ru (Guiraud-Weber 1984): The referent is indeterminate, but delimited by adverbials. Additionally, there is a truly generic reading, as in

(8) Ležaščego   ne b'jut. [Ru]
     lie-CVB.PRS-ACC.SG not beat-3PL
     'One does not beat someone lying on the ground.'

<div align="right">(Guiraud-Weber 1984, cit. after Berger 1991, 74)</div>

Using a system of distinctions advanced by Padučeva (1985) and Berger and Weiss (1987), the former reading can be characterised as [+ref(erential), -def(inite)] (Berger 1991, 75), i.e., there is a referent in the actual world, but it cannot be derived from the context – in more widespread terminology, a specific indefinite. According to the author, there are four possible readings for the 3PL indefinite impersonal in Ru. Simplifying somwehat, these are formed by two binary distinctions, namely, [generic vs. from a closed set] and [new vs. from an inferentially definite set]. (9-a) is an example of a generic use without any inferentially definite set of potential referents around. (9-b) is still generic, but this time the set of potential referents is inferentially definite, i.e., in principle a subset of a definite set (doctors in Čexov's time). (9-c) cannot be viewed as generic, it is a non-specific indefinite, an element of a newly introduced, closed set. In (9-d), the set of potential referents is in principle definite and the event is characteristic for them.

(9) a. Cypljat   po oseni   sčitajut. [Ru]
         chicken-GEN.PL in autumn-DAT count-3PL
         '(lit.) You count the chicken in autumn. (= Don't count your chickens before they hatch.)'   (Kosta 1990, 54, citing Padučeva 1985 and Berger 1991)

     b. Ved' tak, navernoe, vo vremena Čechova uže   ne vračevali. [Ru]
         MP   so   probably   in time-ACC.PL Č.-GEN.SG already not doctor-PT.PL
         'But they probably did not treat patients like this in Čechovs times any more.'

<div align="right">(Berger 1991, 74)</div>

     c. Stučat.   Zovut.   Kričat.   [Ru]
         knock-3PL call-3PL scream-3PL
         'Somebody is knocking / calling / screaming.'   (ibid.)

     d. Arestovali Mandel'štama. [Ru]
         arrest-PT.PL M.-ACC
         'They arrested Mandel'štam.'   (ibid.)

Cz has 3P non-generic, inferentially definite readings akin to (9-d) – in fact, the majority of examples, cf.

(10) Ulici   Veveří už   opravili. [Cz]
      street-ACC V.   already repair-PT.PL
      'They already repaired Veveří street.'   (MČ III, cit. from Berger 1991, 76)

Non-specific indefinite readings (also sometimes called *arbitrary* readings), i.e., direct translations of (9-c), are excluded, and so are purely generic readings like (9-a). Generic readings pertaining to a referentially definite set, as in (9-b), seem to be possible (albeit peripherally), when they describe a behaviour, rather than a norm (Berger 1991, 79):

(11) V Anglii jezdí nalevo. [Cz]
   in England drive-3PL left
   'In England, they drive on the left-hand side.'   (Berger 1991, 79)

It seems that the speaker cannot be included in the set of subject referents in these cases (although Berger does not take a stance on this). Note that the Cz 3PL indefinite impersonal thus shares exactly the realm of uses of English non-referential *they*.

Kosta (1990, 51), following Daneš et al. (1987, 237ff), claims explicitly that in Cz 3PL null pronominal impersonals, the group of subject referents can never include the speaker or the hearer (cf. also Rytel-Kuc 1990). This is understood as the reason for their not being used generically: The predication could not apply to everybody in the universe of discourse. Kosta further cites Weiss (1977, 109) with the finding that also Pol 3PL indefinite impersonals block the inclusion of the speaker. In fact, Weiss' claim is even stronger. He says: *"ferner impliziert die reflexive Variante b) den Einbezug des Sprechers in die Handlung, a) und c) schließen ihn aus." – 'furthermore, the reflexive variant b) [= a Reflexive Impersonal, R.M.] implies the inclusion of the speaker in the action, a) [= a 3PL null subject] and c) [= a -no /-to-form] exclude it'* (Weiss 1977, 109). According to Rytel-Kuc (1990, 144f), Cz and Pol 3PL null subjects mainly differ in relative frequency, but not in their semantic features. We might then want to find constellations which somehow "escape" the ban on generic readings, in a manner similar to (11). Such examples are rare, but they can be found:

(12) [„Pan jest z Mławy". „No tak" – potwierdziłem. – "'You are from Mława". "Well, yes", I confirmed.']

   „U was   biją Cyganów" [– orzekła   pani]   [Pol]
   at you-GEN.PL beat-3PL gypsy-ACC.PL answer-PT.F.SG woman-NOM
   "'At your place, they beat up gypsies" – retorted the woman.'   (NKJP)

(13) [Tak czy inaczej, niebawem na pierwszych stronach gazet przeczytałem: – 'Be that as it may, next I read on the first pages of the newspapers:']

   „W Mławie trują". [Pol]
   in M.-PRP beat-up-3PL
   "'In Mława, they beat up people." ' (Tygodnik Ciechanowski, 02.08.2007 – NKJP)

Both (12) and (13) represent generalising readings, applying to (a subset of) the set delimited by the adverbials *u was* and *w Mławie*, respectively. Thus, it seems that Pol also shares with Cz some of the finer-grained distinctions.

Zolotova (1991, 512) contains some remarks on the clusivity properties of Ru indefinite impersonals. She finds that the speaker is generally excluded from the set of potential subject referents, or at least the construction as such allows for somehow obviating reference to him. In cases of special expressivity (*"govorjat vam ..."* – *'you are being told ...'*), he is understood as included, although actually, reference is not to a certain speaker and a certain speaking event, but carries the connotation of a multitude of speakers and repetitive action. Prescriptive generic uses (as in (1-a) above) are speaker-exclusive, but their communicative purpose may be to strongly influence the addressee and thus, in a way, include him in the set of subject referents. Although Zolotova remains unclear about less prescriptive generic cases, it seems clear to me that the norm expressed in an example like (9-a) is presented as applying also to the speaker. In fact, it is also seems hard to imagine a situation in which someone heavily smoking utters (1-a) – there would be a distinct flavour of self-contradiction.

To sum up, Ru 3ᴘʟ indefinite impersonals allow for generic and indefinite uses, where the set of potential referents may be a subset of a contextually derivable set or not. The truly generic uses and even some non-generic uses may be understood as speaker-inclusive, partly depending on expressity. Cz and Pol 3ᴘʟ indefinite impersonals primarily support the reading of a subset of a known group, designated by adverbials or verbal lexis. To a limited extent, generalising readings related to a (subset of a) known group are possible. The speaker is generally excluded from the set of potential subject referents.

5.2.3 Syntactic properties

In the present subsection, I will merely point out the most important generalisations about Reflexive Passives/Impersonals in Cz and Pol, leaving the full picture of reflexives and also analytical accounts for section 5.4. It is instrumental to first distinguish Reflexive Passives/ Impersonals clearly from so-called *middles.* Middles (as defined in Kaufmann 2004, see below section 5.4.1) obligatorily bring about a modal interpretation, either possibility or an evaluative flavour. An example of a middle (in its most obvious interpretation) is

(14)   provoloka gnetsja   [Ru]
        wire-ɴᴏᴍ   bend-3sɢ-ʀᴇꜰʟ
        'It is possible to bend the wire. (*lit.* The wire bends.)'   (Kaufmann 2004, 205 (44-a))

Ru reflexives formed from *intransitive* verbs always induce a modal interpretation. This may be due to them being middles, or belonging to the so-called *Feel-like construction* (Franks 1995; Marušič and Žaucer 2006) as in (15-a), or *Involuntary State construction*, as in (15-b), each with specific properties:

(15)   a.   Mne   ne rabotaetsja.
              me-DAT NEG work-3SG-REFL
              'I do not feel like working / cannot work well.'   (Franks 1995, 364)

       b.   Rebënku chorošo spalos'.
              child-DAT well   sleep-PT.N.SG
              'The child slept well.'

Involuntary State constructions are also present in Pol (16):

(16)   Tę   ksiąkę   czytało   mi   się z   przyjemnością.
       this-ACC.SG book-ACC.SG read-PT.N.SG me-DAT REFL with pleasure-INS
       'It was a pleasure for me to read the book.'

                                (Rivero and Milojević-Sheppard 2003, 97 (7-a))

Importantly, Reflexive Passives as in (17-a) – i.e., with a NOM patient argument – are generally restricted to middles in modern Pol (cf. the contrast with (17-b)) and obsolete (Siewierska 1988) or even ungrammatical in an episodic context. This is a rather recent development (cf. section 5.5.5). Cz Reflexive Passives show no such restriction (18):

(17)   a.   Ta powieść   przyjemnie się czyta.
              this story-NOM.SG.F enjoyably   REFL reads
              'This story reads well.'   (Szymańska 2000, 29)

       b.   *Piłka rzuciła   się do wody.
              ball-NOM throw-PT.F.SG REFL in   water

                                  (Szymańska 2000, 140)

(18)   a.   Míč   se   hodil   do vody.
              ball-NOM REFL throw-PT.M.SG in   water
              'The ball was thrown into the water.'

       b.   ... způsoby, jakými se stížnosti   vyřizují, ...
               ways   which-INS REFL complaints-NOM manage-3PL
              '... the ways in which complaints are managed'   (ČNK)

Reflexive passives show both hallmarks of canonical passive constructions in European languages: The patient of the corresponding active clauses gets promoted to subject, and the agent of the active clause gets demoted. This sets Reflexive Passives apart from Reflexive Im-

personals (to be discussed below), which only demote the agent of the corresponding active clause, without also promoting its patient. Depending on further considerations, Reflexive Passives and Reflexive Impersonals may be conflated under the heading of *agent demotion* (cf. Chrakovskij 1973, 1974; Comrie 1977; Siewierska 1988; Blevins 2003). The main reason for me to assume two subtypes here is their different diachronic development (cf. section 5.5). Reflexive Passives on the basis of *transitive* verbs can be formed in all Slavonic languages. Major differences concern the admissibility of agent realisation as a *by*-phrase: While East Slavonic languages, Upper Sorbian and Bulgarian allow overt *by*-phrases, Pol, Cz, Slovak, BCS and Slovene do not (Fehrmann et al. 2010), cf. (19) for Ru, Pol[1] and Cz:

(19) a. Dom   stroitsja   (plotnikami).   [Ru]
          house-NOM build-3SG-REFL carpenter-INS.PL
          'The house is being built by carpenters.'   (Fehrmann et al. 2010, 210 (17))

      b. Ta powieść   przyjemnie się czyta (*przez dzieci). [Pol]
          this story-NOM.SG.F enjoyably REFL reads by   children
          'This story reads well (*by children).'   (= modified (18-a))

      c. Šaty   se   právě šijí   (*babičkou).   [Cz]
          dress-NOM.PL REFL just sew-3PL grandmother-INS.SG
          'The dress is just being sowed (*by the grandmother)'
                                                          (Fehrmann et al. 2010, 211 (20))

      d. Èto javlenije   issleduetsja   učënnymi.   [Ru]
          this phenomenon-NOM.SG.N research-3SG.N.REFL scientist-INS.PL
          'This phenomenon is being researched by scientists.'
                                                          (Siewierska 1988, 258 (32b))

      e. Diamantový prsten   se   našel   (*bratrem).   [Cz]
          diamond   ring-NOM.SG.M REFL found-3SG.M brother-INS.SG
          'The diamond ring was found (*by the brother).'   (Siewierska 1988, 259 (35))

Reflexive Passives in this sense include a grammatical subject which agrees with the verb and is marked by nominative case. There is no indication of any further (empty) subject, nor would there be a slot in the structure for one.

A special case of Reflexive Impersonals concerns verbs which optionally take an accusative or a prepositional object, such as verba dicendi, sentiendi, and verbs of propositional attitude (*acc/non-acc* verbs, in Fehrmann et al.'s terminology).

---

[1] to the extent that the Reflexive Passive is acceptable at all in Pol (see above)

(20) Ob   ètom (nami) uže   govorilos'.   [Ru]
      about this we-INS already talk-PT.N.SG.REFL
      'This has already been talked about (by us).'   (Fehrmann et al. 2010, 211)

With respect to the option of *by*-phrase realisation, they behave exactly like Reflexive Passives based on transitives, in all Slavonic languages (Fehrmann et al. 2010, 211f, (22)-(27)) – in Ru, they would allow *by*-phrases, in Cz or Pol they would not.

In Pol and Cz, reflexivisation can also involve *intransitive* verbs, both unaccusatives (21) and unergatives (22). In all of these cases, the realisation of a *by*-phrase is precluded (possibly with some variation in Pol, cf. Siewierska 1988, 258f).

(21) a. Wtedy umierało się na tyfus. [Pol]
          then die-PT.N.SG REFL of typhus
          'At that time one died of typhus.'   (Fehrmann et al. 2010, 213 (34))

      b. Pst, na jevišti se   právě umírá (*hrdinou / *od hrdiny)! [Cz]
          sh at stage REFL just die-3SG hero-INS   by hero-GEN
          'Sh, one is just dying on the stage!'   (Hudousková 2010, 20 (38-a))

(22) a. *Tańczyło   się przez uczniów.   [Pol]
          dance-PT.N.SG REFL by   pupils-ACC.PL
                                                        (Kibort 2006, 198 (42))

      b. Tancovalo se   (*námi) až   do rána.   [Cz]
          dance-PT.N.SG REFL we-INS until to morning-GEN
          'There was dancing (*by us) until the morning.'
                                                        (Fehrmann et al. 2010, 212 (28))

The exact delimitation of the verb classes which allow for Reflexive Impersonals varies somewhat across languages. In Pol, even *być* is a good candidate:

(23) Było się bitym   przez kaprala. [Pol]
      was REFL hit-INS.SG.M by   corporal
      'One was beaten by the corporal.'   (Kibort 2006, 298 (38))

According to Hudousková (2010, 21), Cz Reflexive Impersonals involving *být 'to be'* and *mít 'to have'* are allegedly excluded, because they "do not assign any semantic roles". This cannot be the true story, as the contrast with Pol shows. The reason for the exclusion of *být* and *mít* from Reflexive Impersonals in Cz must be language-specific rather than generally semantic.

A further, more important difference between Pol and Cz concerns syntactic subject properties. Pol, but not Cz Reflexive Impersonals allow for the binding of anaphors and for control into infinitives and gerunds (Růžička 1986b; Rivero and Milojević-Sheppard 2003; Kibort

2004):

(24) a. Teraz się myśli tylko o   sobie.   [Pol]
        now   REFL thinks only about SELF-PRP.SG
        'One only thinks about oneself now.'

        (Rivero and Milojević-Sheppard 2003, 106 (24-a))

     b. Wsiadając   do autobusu pokazuje się bilet.   [Pol]
        get-on-CVB.PRS onto bus   shows-3SG REFL ticket-ACC
        'On getting on the bus one shows the ticket.'   (Kibort 2006, 299 (44))

(25) a. *Mluvilo   se tam jen o   sobě.   [Cz]
        talked-PT.N.SG REFL there only about SELF-PRP.SG

        (Růžička 1986, 255)

     b. *slíbilo   se pomoci dětem   [Cz]
        promised-PT.N.SG REFL help children-DAT

Examples (24)-(25) clearly indicates that Pol, but not Cz Reflexive Impersonals contain a syntactically operative null subject. Furthermore, Pol, but not Cz has Reflexive Impersonals formed from transitives which continue to take and accusative object (so-called Reflexive Accusative Impersonals – Fehrmann et al. 2010, 213):

(26) a. Tu się pije   wódkę.   [Pol]
        here REFL drink-3SG vodka-ACC
        'One drinks vodka here.'   (Rothstein 1993)

     b. *Tady se pije   vodku.   [Cz]
        here REFL drink-3SG vodka-ACC

Due to the observations illustrated in (24)-(26), Reflexive Impersonals are generally viewed as *active* constructions in Pol grammar (Dyła 1983; Siewierska 1988, 263f; Kibort 2004, 289). Usually, they are said to involve an "anonymised" or arbitrary, human null subject (Siewierska 1988, 262). As Kibort (2004) demonstrates, the +HUMAN interpretation is only a default which can be contextually overridden:

(27) Gdy się jest bocianem, gniazdo buduje się wysoko. [Pol]
     if   REFL is stork-INS nest-ACC builds REFL high
     'When one is a stork, one builds the nest high up.'   (Kibort 2004, 292 (431))

Krzek (2010, 77)views this observation in connection with Moltmann's (2006) remark that overt *arbs* are not exactly restricted to human, but rather, to conscious beings (see below). She states that (27), but not (28) is acceptable in Pol; thus, the *arb* null subject has to be +CONSCIOUS.

(28) ??Jeśli jest się szafą, stoi się w kącie. [Pol]
      if be-3sg refl wardrobe-ins stand-3sg refl in corner-prp
      'If one is a wardrobe, one stands in the corner.' (Krzek 2010, 77 (25))

Krzek furthermore claims that the possible readings of Reflexive Passives and Impersonals correlate with verbal aspect: A generic (*"one"*) reading, according to her, is possible only in the imperfective. In the perfective aspect, the reading is arbitrary (*"they/some people excluding the speaker"*), or specific (*"we"*) with transitive and unergative verbs, and unambiguously specific with unaccusatives. These observations are based directly on intuitions about constructed isolated sentences, rather than on acceptability in context, so some caution seems advisable.

Leaving this point for further research, we may summarise that Ru, Cz and Pol lay out a system of fine-grained distinctions among the syntactic properties of reflexivised verbs: Ru allows only transitive verbs to form a Reflexive Passive, other verb classes involving obligatorily modal readings. The Ru Reflexive Passive can express an agentive adjunct in ins case. Cz forms Reflexive Passives and Impersonals from transitive and intransitive verbs, disallowing agentive adjuncts competely, and showing no binding or control by the null subject in Reflexive Impersonals. Nevertheless, there is a restriction to humans (and, possibly, animals). Pol forms Reflexive Passives and Impersonals from transitives, intransitives and even *być 'to be'*, disallowing agentive adjuncts and implying a restriction to humans (and animals). At the same time, the null subject in Reflexive Impersonals behaves like a fully active syntactic subject, i.e., binds reflexives and controls into gerunds and inifinitivals. We will turn to theoretical accounts of the properties reported here in section 5.4.

5.3 Indefinite impersonals in a general linguistic view

5.3.1 Overt Indefinite Impersonals

Impersonal subjects of the kind to be discussed here have *overt* manifestations in many langages. As a base for comparison, let us first concentrate on these overt instances. Well-known examples are e.g. German *man*, English *one*, and French *on*, cf.

(29) a. Man sieht nur mit dem Herzen gut.
          man sees only with the heart well
          'One sees well only with the heart.' (St.-Exupéry, cit. Malamud to appear b, 2)
      b. John doubts that one can see the picture from the entrance.

(Moltmann 2006, 259 (3))

c. On doit travailler jusqu'a l'age de 65 ans.
     one must work   until   the age of 65 years
     'One has to / We have to work until the age of 65.'

(Sigurðsson and Egerland 2009, 162 (7-a))

All of the impersonal indefinite markers (or *arbs*, as Cinque 1988 calls them) in (29) support a *generic* reading, i.e., abstract away from the given concrete event or state. – Authors differ somewhat in their use of the term *generic*. Sigurðsson and Egerland (2009) paraphrase their concept of a *generic null subject* as "like generic English *you* (and generic *one*, in more formal registers)" (p. 158), a "non-restricted +human reading, i.e., people in general", which potentially includes also speaker and hearer (p. 161). This property is brought to light in continuations as in (30):

(30) a. To find the station you/one first turn(s) to the right, *or at least I always do.*

     b. They are on strike in the hotel, *#or at least I am.*

(Sigurðsson and Egerland 2009, 162 (8)/(9))

(30-a) is generic, while (30-b) represents an *arbitrary* reading, roughly synonymous to "someone, not speaker or hearer". Incidentally, some lexical indefinites allow a generic, but not an arbitrary reading – in other words, speaker and hearer must be included in the set of subject referents. These include English *one*, Icelandic *maður*, Hungarian *az ember*, BCS *čovjek* and Sln colloquial *člouk* (*človek*) (Sigurðsson and Egerland 2009, 162, fn.7). *Maður*, notably, can also be interpreted specifically, i.e. as referring to a certain group including speaker and/or hearer.

A generic reading requires unspecified time reference or rather, a lack of delimitation, as well as imperfective aspect (p. 172). Following D'Alessandro's (2006) analysis of these facts in Italian *si*-constructions, aspectual delimitation necessarily leads to a non-generic, *specific* reading, synonymous to *we* in English (cf. (31) vs. (32) from D'Alessandro and Alexiadou 2002, 35):

(31) In quel ristorante si è mangiato bene. [It]
      in that restaurant SI is eaten-PF   well
      'We ate well in that restaurant'   (inclusive)

Lack of boundedness leads to a generic reading *excluding* speaker and hearer:

(32) In quel ristorante si mangiava bene. [It]
      in that restaurant SI ate-IMPF   well
      'One used to eat well in that restaurant'   (generic)

Note that speaker inclusion is automatically coupled with non-genericity here. That may be right for canonical null subject languages, but – as Holmberg (2010) shows, see below – it does not carry over to partial null subject languages.

D'Alessandro's (2006) claim partly contradicts Chierchia's (1995) seminal analysis, so we will recapitulate the main points of the latter here. Chierchia (1995) offers the first explicit semantic treatment of Italian *si*. In generic sentences, the *si* subject gets a quasi universal reading (roughly paraphrasable as *all people* – (33-a)), while episodic sentences favour an existential reading (*someone/some people* – (33-b)).

(33)   a.   In Italia si beve   molto vino [It]
              in I.   si drink-3sg much   wine
              'In Italy, everybody/people drink(s) a lot of wine'

        b.   In Italia ieri   si è   giocato male [It]
              in I.   yesterday si aux-3sg play-cvb poorly
              'Yesterday, people in Italy played poorly'   (Chierchia 1995, 108)

More precisely, (33-a) means that in Italy, in many contextually relevant situations, there is drinking going on. Unless a contextually relevant situation is available, generic sentences will receive *no* interpretation (Chierchia 1995, 123). Chierchia does not explicitly address the question whether the speaker and/or hearer should be exempt from the group of subject referents in generic sentences (as claimed by D'Alessandro 2006), but he does not assume any such restriction for *episodic* sentences (p. 122).

Chierchia's analysis centers around the observation that *si* shows so-called *quantificational variability effects* (QVEs), a hallmark of true indefinites. In Italian, the effect also holds for bare plural NPs, although not in a completely parallel fashion as for impersonal indefinite pronouns (cf. Chierchia 1995, 108). *Si* shares with indefinites the property that a quantificational adverbial in a main clause can restrict the quantificational force of an indefinite in a subordinate conditional *if*-clause, as in (34):

(34)   Se si è   alti, si è   talvolta   anche belli   [It]
        if si aux tall si aux sometimes also   beautiful
        'Some people who are tall are also handsome.'   (Chierchia 1995, 110–(9))

Si can itself bind, i.e., act as an antecedent of reflexives (35-a), and control into infinitives (35-b). However, it cannot act as the antecedent of pronominals or definite *pro* (35-c), only of a further impersonal *si* (35-d) (= Chierchia 1995, 109 (5)-(8)).

(35)   a.   Ci   si è   lavati.
              self-cl si aux washed

'People washed themselves.'

b. si$_j$ è cercato di [PRO$_j$ vincere]
    people AUX-PAST tried to win

c. si$_j$ è detto che pro$_j$ vinceranno
    SI AUX-PAST said that will-win

d. Ieri, si è giocato male e si è perso.

'Yesterday, people played badly and they/people lost.'

The gist of Chierchia's analysis is to treat *si* itself as indefinite. In particular, he proposes that *si* takes a property, existentially closes the argument corresponding to the subject, and introduces a restriction of this argument to groups of humans (Chierchia 1995, 120). Formally, a sentence like (36-a) then becomes analysed as (36-b):

(36)   a.   si canta
             people sing
        b. $\exists x_{arb}$ [sing (x_{arb})], with x_{arb} ranging over groups of humans

This seems adequate enough for episodic sentences. *Generic* sentences contain a generic operator Gn which could either take maximal scope or only apply to the VP, remaining within the scope of the existential quantifier which binds x_{arb}. In the first case, Gn can only bind the event variable, which for (36) would result in an interpretation paraphrasable as "In every contextually relevant situation, people sing". In the latter case, the reading would be something like "There is a (salient) group of humans which have the habit of singing". According to Chierchia, both generic readings are possible.

In passing, Chierchia makes several observations which turn out as extremely important for the present work and the issue of null subjecthood in general. Discussing the *episodic* reading of *si*, he notes that (37-a) clearly excludes the speaker from the group of subject referents, while (37-b) favours a speaker-oriented interpretation (Chierchia 1995, 126):

(37)   a.   *pro* ti hanno cercato
             pro YOU-CL AUX-PAST look-for-PAST
             'somebody was looking for you'

        b.   ti si è cercato
             you-ACC-CL SI AUX-PAST look-for
             'People/we looked for you'   (Chierchia 1995, 126 (35)/(36))

This observation for Italian is exactly in line with Holmberg's (2010) claim that canonical NSL do not have pronominal null subjects with a generic, speaker-*inclusive* reading, in active clauses (see below). Chierchia is, of course, not concerned with the pro-drop debate, but

instead focuses on the semantic consequences of the difference in (37), arguing that it forms the reason for which *si* and *pro*-3PL cannot be anaphorically linked to each other. *Si*, by assumption, introduces a special index on its variable, restricted to groups of humans, which cannot be coindexed with another bound variable. While this is rather clear in the case of definite *pro* vs. impersonal *si*, it is a bit more involved in the case of binding of reflexives and control into infinitives. That *si* can bind reflexives is to be expected under an analysis in which there is no external binding of two independent variables, but a reflexive property is predicated of *si*, which Chierchia indeed proposes (p. 127). In a similar spirit, he argues that embedded infinitives can be analysed as properties, which are predicated about *si*, rather than propositions with a null argument. So again, *si* need not bind (which it is unable to), but instead, becomes the subject of a "self-oriented" predicate. Cases in which *si* seems to be coreferent with another *si*, do not involve binding either, following Chierchia. Instead, they are linked by pragmatic mechanisms. Finally, quantificational adverbs have the general ability to disclose existential quantification over indefinites and bind off the resulting variables again. Nothing else has to be stipulated for quantificational variability effects with *si* in conditional clauses.

A second remark, delegated to a footnote (Chierchia 1995, 141, fn.7), sheds even more light on the semantics of *si* vs. *pro*-3PL: While *si* shows QVEs (cf. (34)), *pro*-3PL does not:

(38) se *pro* sono alti, *pro* sono sempre / di solito / raramente anche belli
      if PRO are tall are always usually rarely also handsome
      'If they are tall, they are always / usually / rarely also handsome.'

As Chierchia comments, the *pro* subject of (38) is not an indefinite, but must refer to a specific group of people or a specific kind.

This type of evidence is taken up again by Malamud (2006; 2007; to appear a; to appear b), who – importantly for us – also includes modern Russian 3PL impersonals and Reflexive Passives in her analysis. *Overt* arbitrary pronouns like *man* or *si* actually show a dual behaviour: Like indefinites, they are subject to QVEs:

(39) a. Damals lebte man normalerweise / selten bis 60 Jahre.
          then lived MAN usually rarely till 60 years
          'In those days, one usually/rarely lived till 60.' (QVE available: most/few people in those days lived to be 60)

      b. Qui si è di solito / raramente tifosi dell'Yankees.
          here SI is usually rarely fans of-the-Yankees
          'Here one is usually/rarely a fan of the Yankees.' (QVE available: most/few people here are fans of the Yankees)

  c. In those days, you usually/rarely lived to be 60. (QVE available: most/few peo-
    ple in those days lived to be 60)   (Malamud to appear a)

At the same time, they can involve an indexical, speaker-including interpretation:

(40) Es war völlig klar, dass man sich nie mehr wiedersehen würde.
   it was completely clear that MAN SELF never again see.again would
   'It was completely clear that we would never see each other again.' (Kratzer 1997, 5)

Typical indexicals (like *I, these students*), however, never show QVEs (Malamud 2007, 8).
Malamud's (2007) formal proposal, partially modelled after Kratzer (1997), makes exclusive
man/si semantically equivalent to the indefinite *a person*. Inclusive *man/si* specifies that the
speaker has to be a subpart of the variable standing for the subject. According to Malamud
(to appear b, 11), there is considerable dialectal (or probably rather idiolectal) variation with
respect to German *man*: Some of her informants only accept speaker-exclusive *man*, others
use inclusive *man* only in non-episodic (= generic) contexts, and a few allow inclusive *man*
in generic, but also in episodic sentences like (40) and (41).

(41) Es ist eine Ewigkeit her, dass man sich nicht gesehen hat.
   it is an eternity ago that MAN each.other not seen has
   'It has been an eternity since we saw each other.'

              (Mann: Mephisto, cit. Malamud to appear b)

There are further interpretations of impersonal indefinite pronouns, differing somewhat
between languages. Sigurðsson and Egerland (2009, 161f) mention an *arbitrary*, non-specific
+HUMAN reading excluding the speaker or the hearer, and a *specific* +HUMAN reading which
mostly includes the speaker, in French. (42) is ambiguous between the latter two.

(42) On a travaillé deux mois pour résoudre le problème.
   one has worked two months to resolve the problem
   'They (arbitrary)/We (specific) worked for two months to resolve the problem.'

          (Sigurðsson and Egerland 2009, 161f)

The fact that (42) disallows a generic reading (and allows a specific, inclusive one) is related
to the presence of the delimiting temporal adverbial (Sigurðsson and Egerland 2009, 173;
cf. also D'Alessandro 2006, 67ff for similar observations concerning Italian *si* and Krzek
and Kibort for Polish impersonals). German *man* and English *one* do not seem to allow
for the specific ("*we*") reading, whereas French *on* and Italian *si* do so. Turning to Slavonic
languages now, Malamud (2006; to appear a), found that Ru 3PL indefinite impersonals do not

show quantificational variability effects with quantificational adverbs at all, and thus differ strongly from English *one* as well as from German *man* and Italian *si.* First note that in Ru, as in English, Italian or German, quantificational adverbs like *usually* and its equivalents *obyčno, gewöhnlich, di solito* induce a QVE with plural indefinites, but not with plural definites:

(43) a. V MGU studenty obyčno xorošie rebjata / ljubjat dekana.
          In MGU students usually good-PL guys / love-3PL dean-ACC
          'Students of Moscow University usually are good people / love the dean' (QVE

          available: Most students = QVE on students)

      b. Èti   studenty   obyčno ljubjat dekana.
          these-NOM students-NOM usually love-3PL dean-ACC
          'These students usually love the dean.' (no QVE: the only reading is 'Now they

          love the dean, now they don't')   (Malamud 2007)

Adverbials like *for the most part* or Ru *bol'šej čast'ju*, on the other hand, *can* induce a QVE on plural definites, cf.

(44) Èti   studenty   bol'šej čast'ju ljubjat dekana.
      these-NOM students-NOM most-INS part-INS love-3PL dean-ACC
      'These students for the most part love the dean.' (QVE on students) (Malamud 2007)

As Malamud goes on to show, Ru 3PL null subjects pattern with plural *definites* as regards QVE, while German *man* or Italian *si* pattern with plural indefinites. I.e., (45-a), but not (45-b) shows a QVE; (45-b) could be paraphrased as something like *Now they love the dean, now they don't*:

(45) a. Na ètom fakul'tete bol'šej čast'ju ljubjat dekana.
          on this   faculty   most-INS part-INS love-3PL dean-ACC
          'In this department, they for the most part love the dean.'

      b. Na ètom fakul'tete obyčno ljubjat dekana.
          on this   faculty   usually love-3PL dean-ACC
          'In this department, they usually love the dean.' (Malamud to appear a, (18c,d))

The observation also extends to Ru reflexive passives:

(46) a. V Rossii, ``Pravda'' bol'šej čast'ju preziraetsja.
          in Russia   Pravda-NOM most-INS part-INS despise-3SG.REFL
          'In Russia, "Pravda" is for the most part despised.'

      b. V Rossii, ``Pravda'' obyčno preziraetsja.
          in Russia   Pravda-NOM usually despise-3SG.REFL
          'In Russia, "Pravda" is usually despised.'   (Malamud to appear a, (20))

The analysis of Ru 3PL null subjects as plural definites has further ramifications concerning their behaviour in discourse. As Malamud argues, they can be taken up by intersentential anaphora (while still being rather low in salience), cf. the constructed example in (47):

(47)   Ivan   uže   privyk, čto ego rasskazu ne verjat   ili ponimajut   ego
        Ivan already got.used that his   story-ACC not believe-3PL or understand-3PL it-ACC
        kak-to   izvraščenno. Oni ved' ne vstrečalis' s   d'javolom.
        somehow perversely   they indeed not have.met   with devil
        'Ivan already got used to the fact that his story was not believed or understood somehow perversely. Unlike him, they haven't met the devil.'   (Malamud 2007 (39))

By contrast, German *man* or Italian *si* can never serve as discourse antecedents for pronouns (other than further *man* and *si* – Chierchia 1995; Kratzer 1997). While 3PL null subjects decrease the prominence of the logical subject, *man* or *si*-type arbitrary pronouns do not contribute to the salience/topichood computation at all, according to Malamud (to appear b). Note that *generically* interpreted *overt* NPs may be taken up by a 3P pronoun in Ru as well, as has been noted by Padučeva (1985, 98).

To sum up, there are further fine-graned semantic distinctions in the readings of arbitrary pronouns across languages. Italian *si* has been shown to support a truly indefinite reading, subject to quantificational variability effects. The availability of such a reading vs. a generic one partly depends on aspectual delimitation. There is some indication that the Ru 3PL null subjects do not behave like a true indefinite, but rather like a plural definite. In this respect, they seem to differ from the null subject present in Reflexive Passives/Impersonals in canonical null subject languages like Italian.

5.3.2 Pro$_{arb}$ and pro drop

Ru has diachronically developed from a consistent NSL into a partial NSL, as discussed in detail in chapter 3. At the same time, it makes ample use of *indefinite* null subjects with 2SG or 3PL agreement, traditionally called *obobščenno-ličnaja forma 'generalising-personal'* and *neopredelenno-ličnaja forma 'indefinite-personal form'*. Cz, on the other hand, which has remained a consistent NSL throughout its history, disallows at least part of the indefinite null subjects of the Ru type.

Holmberg (2010) argues that the pattern we are observing here is, in fact, very widespread and should be encoded in the grammar of partial vs. consistent null subject languages (NSLs). Those languages which consistently allow definite[2] null subjects, such as Italian or European

---

[2] We have been using the term *(co)referential* in place of *definite* so far.

Portuguese, have no null generic subject pronoun (48-a). *Partial* NSLs, like Brasilian Portuguese or Finnish, do allow for non-definite (generic) null subjects, cf. (48-b). Non-NSLs, such as English or Swedish, of course disallow any null subjects (apart from "diary drop" and related cases).

(48)   a.   É assim que se faz   o   doce. [European Portuguese]
            is thus   that SE makes the sweet
            'This is how one makes the dessert.'

       b.   É assim que faz   o   doce. [Brasilian Portuguese]
            is thus   that makes the sweet
            'This is how one makes the dessert.'   (Holmberg 2010, 92)

It must be noted that this generalisation is intended to apply only to null subjects with certain readings, namely those *including* the speaker and the addressee (= English *one*), rather than to those *excluding* them (English generic *they*). The *exclusive* generic reading can appear in consistent NSLs for 3PL null subjects.[3]

Of the partial null-subject languages, some accept null exclusive as well as inclusive generic pronouns (Brasilian Portuguese), others just accept inclusive readings (Finnish). Note that Holmberg's "generic exclusive" reading is exactly the one described above as the generalising reading applying to an open subset of an inferentially definite set (Berger 1991), which could also be found in Cz and Pol, albeit peripherally. The respective examples are repeated here for convenience:

(49)   V Anglii jezdí   nalevo.
       in England drive-3PL left
       'In England, they drive on the left-hand side.'   (Berger 1991, 79)

(50)   [Tak czy inaczej, niebawem na pierwszych stronach gazet przeczytałem: – 'Be that
       as it may, next I read on the first pages of the newspapers:']

       „W Mławie trują".
       in M.-PRP beat-up-3PL
       "'In Mława, they beat up people." '   (Tygodnik Ciechanowski, 02.08.2007 – NKJP)

Recall that the "true" generic reading, expressing a generalisation or norm without any related referent set in the discourse, and including the speaker (*contra* Zolotova 1991), was impossible in Cz or Pol, but well-attested in Ru.

---

[3]Holmberg's examples from partial NSLs are all 3SG (although he mentions Ru, with its generic 3PL null subject, under the same rubric). Otherwise the issue of number is not treated in detail. Furthermore, the generalisations only apply to active clauses; passives or impersonal passives have to be treated separately (cf. also Fassi Fehri 2009, discussed below).

Holmberg's implicational rules may be summarised as follows:

(51)   a.   consistently referential pro ⇔ no generic speaker-inclusive 3p null subject

         b.   partially referential pro ⇔ generic speaker-inclusive 3p null subject

         c.   no referential pro ⇔ no generic 3p null subject

Sigurðsson and Egerland (2009) criticise Holmberg's conclusions on the following grounds: (i) Icelandic, a non-NSL, has developed null subjects in certain impersonal constructions with a generic reading, contrary to (51-c) (or to (51-b), if it was a partial NSL – for which there is, however, no evidence (Sigurðsson and Egerland 2009, 180)). The relevant constructions are the impersonal passive, an impersonal present participle construction, and an impersonal modal construction. (ii) It is unclear why (51) pertains only to generic, rather than also to arbitrary *pro*. (iii) Old Norse had both definite and generic pro. (iv) The typology has to be more complicated in two respects: One may introduce more relevant factors, such as the existence of an overt impersonal subject (German *man 'one'*, Ru *čelověk* , Cz *člověk 'man/human being'*) or of *specific* readings. Moreover, there is no reason – according to the authors – to exclude any of the logically possible combinations of generic and definite null subjects. – The empirical content of arguments (i) and (iii) has to be evaluated. Regarding Icelandic, it is suspicious that generic null impersonals are restricted only to those three constructions. The Old Norse evidence would have to be scrutinised diachronically, show that it was not just a stage of transition. (ii) and the first part of (iv) call for extensions to Holmberg's theory, without ultimately refuting it. The second part of (iv) is just a claim *contra* Holmberg, supported by no evidence. Thus, to the extent that satisfying explanations for the partially relevant Icelandic and Old Norse facts can be found, Sigurðsson and Egerland (2009) present no real counterargument against Holmberg's analysis.

We may add that Chierchia (1995), when observing the difference in clusivity between *si* and *pro*-3PL in Italian (see above), explicitly discussed the *episodic* uses. So to the extent that Holmberg's generalisation concerning canonical NSLs is correct, it may carry over to episodic (arbitrary, existential) readings as well: They may be blocked in canonical NSLs, but exist in partial NSLs. Ru, Pol, and Cz would fit this picture.

The technical implementation of Holmberg's findings runs as follows: Holmberg (2010, 94f) proposes – following partly Holmberg (2005) – that the unique property of *consistent NSLs* is the presence of a D(efinite)-feature on finite T, as part of its φ-features. Pronouns, in general, are either DPs, with an internal structure as in (52), or φPs; null pronouns are always φPs.

(52) $[_{DP}$ D $[_{\varphi P}$ φ $[_{NP}$ N $]]]$ (Holmberg 2010, 94)

T comes with the unvalued φ-features number, gender and person. By the operation of agreement, T receives values for these features and simultaneously transfers its D feature to the null subject, which becomes a definite pronoun. If D is lacking on T, as Holmberg assumes for partial NSLs, the null subject remains D-less, thus non-definite. The interpretation is that of an impersonal, i.e., generic or non-thematic; in the case of 3SG/PL, a generic inclusive interpretation obtains. Existential indefinite readings are impossible due to the lack of a quantificational component in the pronoun. Holmberg partly adopts Frascarelli's analysis of the relation between Aboutness-shift topics and pro licensing. Specifically, he assumes that the referential index on D in T is contributed by a phonetically empty Aboutness-shift topic in the left periphery of the clause, whose index in turn arises from identification with the closest overt Aboutness-shift Topic in the discourse. Within the adopted framework, the index transferred to T adds a definite interpretation and provides a value for T's unvalued D feature.[4] Agreement between T and the null subject then gives rise to a chain of which only the highest member is pronounced, i.e., the affix on the finite verb or auxiliary. In the case of an overt subject DP, which usually represents a new Aboutness-shift topic itself, agreement holds between T and this DP.

In a *partial NSL*, following Holmberg, T lacks a D feature completely, and thus agreement with a null subject pronoun cannot yield a definite reading. If the interpretation is that of a coreferential pronoun, the null subject has to be controlled by a local antecedent. This control relation is analysed in more detail by Holmberg and Sheehan (2010).

The distribution of generic vs. coreferential readings in *partial* NSLs is further restricted by word order in an interesting way in Finnish, cf. (53-a,b):

(53)   a.   Jari sanoo että tässä istuu mukavasti.
               Jari says   that here   sits   comfortably
               'Jari says that one can sit comfortably here.'

         b.   Jari sanoo että Ø istuu mukavasti tässä
               Jari says   that   sits   comfortably here
               'Jari says that he sits comfortably here.'   (Holmberg 2010, 102 (21))

(53-a), with the local adverbial in preverbal position, only allows for a generic reading, whereas (53-b) is only definite (coreferential). Holmberg argues that subjects have the option of either incorporating into T (by the Agree relation discussed above) or moving to SpecTP.[5]

---

[4]The compositional semantic details of this analysis are unfortunately left aside by Holmberg (2010).

[5]Technically, both processes are related to the influence of the EPP. As Holmberg assumes, the EPP is satisfied by the empty Aboutness-shift topic in the former case, without SpecTP being projected at all; in the latter

Since the SpecTP position is already filled by the local adverbial in (53-a), the null subject incorporates into T and, due to lack of D in T, is interpreted generically. In (53-b), on the other hand, the null subject fails to incorporate into T and moves up to SpecTP. By assumption, it already comes with a D feature and gets it valued in SpecTP, via a control relation with a locally nearby DP using an 'extended version of chain reduction' (Holmberg 2010, 104). Holmberg continues with an in-depth discussion of non-NSLs, into which we cannot embark here. The special influence of locative phrases has also been pointed out by Cabredo Hofherr (2003; 2006), who discusses the fact that locative, but not temporal adverbial phrases can serve as anchors to establish a "corporate existential" reading excluding speaker and hearer in Spanish:

(54)   a.   Durante el servicio militar aprenden a utilizar un arma. [Spanish]
              'During the military service (they) learn to use a gun.' (non-anaphoric ?*)
        b.   En la mili aprenden a utilizar un arma. [Spanish]
              'In the military (they) learn to use a gun.' (non-anaphoric ok)

<div align="right">Cabredo Hofherr 2003, 92 (31b,c)</div>

Cabredo Hofherr herself hypothesizes that this "may be due to the fact that locatives share properties with nominal subjects" – this is exactly Holmberg's argument.

As acknowledged by Holmberg (2010, 93), there is very clear evidence for an important qualification and refinement of his implications between the null subject property and the existence of generic readings, which has been put forward recently by Fassi Fehri (2009): Consistent NSLs only display "truly" generic null subjects in constructions with *passive or impersonal voice*. If this is to be modelled syntactically, the relevant feature specifications of *pro* must be somehow related to the derivation of the passive and the impersonal. Fassi Fehri (2009) mainly argues on the basis of (Standard) Arabic, but also points to corresponding evidence from Finnish, Irish, and Italian in the literature. Arabic is a consistent NSL, which also allows for generic null subjects in constructions with a passive verb:

(55)   y-u-jlas-u   hunaa   waqt-a   l-istiraaḥat-i
        3.PAS-sit-IND   here   time-ACC   the-break-GEN
        'One sits here at break time.'   (Fassi Fehri 2009, 7)

The passive/impersonal illustrated in (55) disallows *by*-phrases, but may contain a direct object, if it is formed from a transitive verb. This is the same array of facts which we find in modern Pol (see above, section 5.2.3). Likewise, Arabic allows for 2SG, 1PL, AND 3PL im-

---

case, it is satisfied by the subject moving into SpecTP.

personal null subjects, the latter only in a sense excluding the speaker and the hearer. Fassi Fehri characterises generic *pro* as [Ø Pers], i.e., as containing a zero person feature, which is sematically restricted to [+human], notably the default range ascribed to the category of person (Cardinaletti and Starke 1999). The minimal structure of a clause with definite *pro* would be as in 5.1.
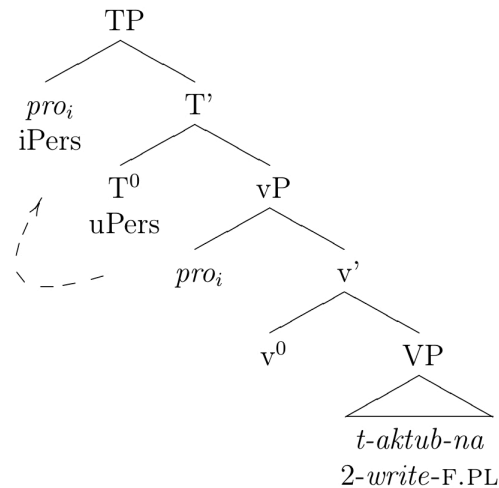

Figure 5.1: Fassi-Fehri (2009, 18 (57))

Pro originates in Spec-*v* like any other external argument, and then moves up to Spec-T to value the unvalued [Pers]-feature on T. As for diathesis, it is instrumental for Fassi Fehri to unite impersonals and proper passives under the aspect of subject demotion, i.e., "a weakening of the 'referentiality' of the subject pronoun, which is encoded on Voice inflection" (Fassi Fehri 2009, 21). The generic *pro* present in impersonals is then specified [ØPers], whereas the demoted argument in passives lacks any such specification. For reasons which are not discussed in detail, Fassi Fehri assumes a Voice phrase (VoiP) between TP and *v*P in the case of reflexive or passive diathesis, whose head contains a [uPers] feature. In a manner exactly parallel to definite *pro* in (55) (only at the VoiP level), *pro* [ØPers] then moves up to Spec-Voi and values the Person feature of Voi[0]. In passives, on the other hand, the external argument lacks the respective Person specification, thus the object moves to Spec-Voi (is probed by Voi), again for reasons of Person marking. The machinery invoked here makes the right predictions, but in a somewhat technical way. Details of the implementation aside, the fact that the generalisations developed by both Holmberg and Fassi Fehri go through in such an array of partially unrelated languages (e.g., Finnish, Italian, Portuguese, Hebrew, Arabic, Irish) is

stunning all by itself.

5.3.3 Predictions for diachrony

If Holmberg (2010) and Fassi Fehri (2009) are on the right track, and their findings carry over also to the realm of diachrony, then we would expect (all else being equal) that changes from a consistent to a partial pro-drop language also bring about a concomitant change for 3PL indefinite impersonals: They should begin to support a generic, speaker-inclusive, non-delimited (= not inferentially definite) reading. On the other hand, a consistent null subject language should not allow for such readings with 3PL null subjects, but only under changes of voice (passive or impersonal).

In our case, the first question would be whether there were any changes in the semantics of indefinite impersonals in the history of Ru, and if so, at what time period these constructions began to support a generic, speaker-inclusive reading. The hypothesis would be that this happened concomitantly with the loss of canonical pro-drop and the rise of weak subject pronouns.

Second, we must check what readings were available for early 3PL indefinite impersonals in Cz and Pol. Here, we would be led to think that there was no change in readings, since both were canonical null subject languages from the beginning. However, this case is also interesting, because, as we will see, Reflexive Passives and Impersonals – i.e., those changes in voice which are claimed decisive by Fassi Fehri (2009) – developed only later.

5.4 Reflexivity and diathesis

5.4.1 Reflexive marking: argument structural distinctions

Geniušienė (1987) provides a throrough overview of the various meanings of reflexive markers, including also Russian examples. Her taxonomy is taken over and slightly extended by Kaufmann (2004); both authors strive for a unified semantic analysis of the various reflexive meanings. Geniušienė argues that diathesis involves mappings between the three analytical levels of (i) referent structure, (ii) semantic roles, and (iii) syntactic structure. Each level has its own hierarchy of constituents. In (i), this hierarchy consists of numbered, sortally restricted referents. (ii) represents an ordering of semantic roles like Agent, Actor, Patient, and others, which may be grouped into the "hyperroles" of Semantic Subject, Semantic Object, and Semantic Dative. (iii) is a hierarchy of grammatical functions (subject, direct object, indirect object etc.). Basic verbs display a parallel ordering with a one-to-one correspondence

of constituents on all three levels. Reflexive verbs involve various modifications of this mapping, of which those affecting the Semantic Subject are most relevant in our context.

The idea that reflexivisation marks a mismatch between different hierarchies (argument-structural and/or syntactic) is developed further by Kaufmann (2004) in her explicit account of the medium voice[6]. Kaufmann couches her analysis in the framework of two-level semantics (Bierwisch 1983, 1997; cf. also Fehrmann, Junghanns and Lenertová 2010). Empirically, she investigates the Niger-Congo language Fula along with Classical, Koiné and Modern Greek, and integrates facts about Germanic and Romance languages and Russian. Her findings concerning the diachronic development of Greek are instrumental for us, in order to assess the hypothesis that SE-passives were borrowed from New Testament Greek (= Middle Greek, Koiné Greek) into OCS (see 5.5.2).

Turning to reflexivisation, Kaufmann (2004, 228) argues that it often fulfils the same functions as a dedicated medium voice marker in other languages, and some more non-medial ones. She distinguishes the following medium (/reflexive) meanings, illustrated here with her Ru examples (which are partly cited from other sources, cf. the references in Kaufmann 2004, 198-209):

(56) a. *directly reflexive:* Maša odevaetsja. – 'Maša dresses.'

  b. *causative-reflexive:* On razvelsja. – 'He got divorced.'

  c. *indirectly reflexive:* zapasat'sja – 'to provide oneself with something'

  d. *reciprocal:* obnimat'sja – 'to hug'

  e. *decausative:*

   (i) *pseudo-reflexive:* Soldaty vystroilis'. – 'The soldiers formed up.'

   (ii) *pseudo-passive:* Dver' otkrylas'. – 'The door opened.'

  f. *modal:* Provoloka gnetsja. – 'The wire bends.'

  g. *passive:* Èta dver' zakryvalas' nami. – 'This door was closed by us.'

  h. *reflexiva tantum:* sadit'sja – 'to sit down'

This classification as such is certainly ver reminiscent of Geniušienė's; the new theoretical point in Kaufmann's approach, however, may be seen in the thorough analysis of semantic *control properties* as relevant for the coding of medium voice.

Medium voice, according to Kaufmann (2004, 44), generally signals that the actually realised control properties diverge from the canonical ones for a given verb. She distinguishes three different kinds of control: (i) S(ituational)-control: An individual *controls a situation*

---

[6]I will use the term 'medium' in the remainder, because the English label 'middle' is often exclusively identified with the modal reading.

if it determines whether and for how long the situation holds. Whether the subject *intends* to control the situation, is not decisive. (ii) Predicate control: A predicate can be designated as requiring a controller at the level of S(emantic) F(orm). Predicates of this kind are called *activity* predicates; only the highest argument in a SF formula can be a controller. (iii) Sortal control properties: Sortal semantic characteristics of a given individual can enable and/or prototypically require it to be a controller.

In languages distinguishing an active and a medium voice, such as Fula, S-control is encoded grammatically: While active voice represents the unmarked case of the predicate controller being also the S-controller and fulfilling the prototypical sortal control restrictions, medium voice encodes the lack of such a harmonic mapping. Thus, e.g., *media tantum* in typologically absolutely diverse languages like Fula and Ancient Greek mostly come from the same semantic classes (verbs of cognition, emotion and movement). In the case of a self-controlled movement verb like Fula *yotto 'to arrive'* or *jippo 'to get off'*, the predicate is not specified for any restrictions on its highest argument (no predicate control). Nevertheless, the verbs are used agentively, i.e., S-control holds (Kaufmann 2004, 52) – the two hierarchies are not aligned with each other. Conversely, the highest argument of a verb of cognition typically lacks S-control, but the verb in itself is an activity, i.e., predicate control applies. Other media tantum like *barro 'to threaten'* or *toro 'to ask for'* require a human Patient argument (sortal control), which is, however, too low in the SF representation to be a predicate controller. An interesting mismatch between control hierarchies may be observed in *causative reflexives* such as Fula *mooro 'to have one's hair cut'* (vs. *moora 'to give (someone) a haircut'*). Kaufmann analyses them as involving S-control by the causer, because he may begin or end the activity, but predicate control by the causee, who is equal to the agent.

Decausatives can be divided into two subclasses, the so-called *pseudo-passives* and the *pseudo-reflexives* (Kaufmann 2004, 60, following the terminology of Rijksbaron 1984):

(57)   a.   *pseudo-passive:*
            Fula: *maBBo 'to close', Besdo 'to rise'*
            Classical Greek: *dúomai 'to sink (intr.)', sbénnumai 'to cease (fire) (intr.)'*

        b.   *pseudo-reflexive:*
            Fula: suuDo *'to hide (intr.)', moBBto 'to gather (intr.)'*
            Classical Greek: *klínomai 'to lean against (intr.), trépomai 'to turn around (intr.)'*

The pseudo-passive reading is the one commonly termed *decausative.* Obviously, no causer

("canonical controller") or S-controller is part of its semantic representation at all. In a pseudo-reflexive reading, the S-controller is the animate Patient argument, which is marked as a potential controller in the verb's lexical entry (as in the directly reflexive reading). However, the canonical controller is not present in the θ-grid (Kaufmann 2004, 61). Note that either aspect of pseudo-reflexive meaning (S-control by the Patient or lack of a controller in the θ-grid) could lead to medium voice marking. (58) and (59) are Kaufmann's (2004, 109 (26)-(27)) semantic representations of pseudo-passive and pseudo-reflexive readings, respectively:

(58) cl. Greek *tḗk- 'to melt'*:

    a.   active (transitive):

       SF: $\lambda y \; \underset{\text{CNTR}}{\lambda x^{<+cntr>}} \; \lambda s \; \{ \; \text{MANIP}_{\text{act}}(x,y) \; \& \; \text{MELT}(y) \; \} \; (s)$

    b.   medium, pseudo-passive decausative:

       SF: $\underset{\text{CNTR}}{\lambda y^{<-cntr>}} \; \lambda s \; \{ \; \text{MANIP}_{\text{act}}(x,y) \; \} \; (s') \; \& \; \{ \; \text{MELT}(y) \; \} \; (s)$

       CS: $\exists \, s', x$

(59) two possibilities for pseudo-reflexive decausatives:

    $\text{SF}_1\text{:} \; \underset{\text{CNTR}}{\lambda y^{<+cntr>}} \; \lambda s \; \exists \; s' \; \{\text{MANIP}_{act} \; (x,y) \; \} \; (s') \; \& \; \{\text{PRED}(y^{pk})\} \; (s)$

    $\text{CS}_1\text{:} \exists \, s', x$

    $\text{SF}_2\text{:} \; \underset{\text{CNTR}}{\lambda y^{<+cntr>}} \; \lambda s \; \{\text{MANIP}_{act} \; (x,y) \; \& \; \text{PRED}(y^{pk})\} \; (s)$

    $\text{CS}_2\text{:} x{=}y$

(58) expresses Kaufmann's generalisation: The use of the medium is due to the mismatch between the predicate requiring a controller (formalised as an annotation CNTR below the lambda binder of the respective argument) and the lack of control on behalf of the non-agent argument (marked as a superscript <±CNTR>). At C(onceptual) S(tructure), there is an implicature that such an argument exists in the (partial) situation.

(59) provides two alternative possible interpretations for pseudo-reflexive decausatives: Note first that both keep the control requirement by the predicate and the controlling property of the respective argument in concord – the subject argument does not lose situational control here. However, the lower argument y is a *potential* controller here (superscript *pk*). (59-a) expresses the parallel with the decausative pseudo-passive reading (58-b), which is to be found in the exclusion of the controlling situation *s'* from the meaning of the predicate; (59-b) only represents S-control by the Patient argument. Kaufmann discusses both options

for Classical Greek, finding evidence for either.

Interestingly, *passive* readings are – by intention – not covered by Kaufmann's generalisation on medium forms. In a passive, S-control is exerted by a non-subject argument. However, this argument is only syntactically 'demoted', remaining the highest one in the SF representation. It may also freely be subject to predicate control and sortal control. The control hierarchies are thus in concord, and there is no motivation whatsoever for the use of medium voice. Following Kaufmann 2004, 98 (57), the agent in a passive structure is part of the θ-grid, but existentially bound off at SF; its identification with an oblique adjunct phrase must be achieved at C(onceptual)S(tructure). While this separation of medium and passive is adequate for Fula, it seems problematic with respect to many Indo-European languages, in which medium forms often acquired passive readings through diachronic changes. On Kaufmann's view, an original medium marker would have to substantially change its semantics in order to become a marker of passive voice; in the long run, it will then either cease to designate medium meanings, or the medium meanings it used to mark begin to erode altogether. Kaufmann mainly addresses this point with respect to Greek, but later extends the analysis to Romance and Germanic languages and Russian.

She manages to compile a comprehensive account of the complicated development of medium and passive readings in the history of Greek, relying on the standard sourcebooks, as well as more specialised studies. In the following simplified overview, I use the terms *medium* and *passive* to name grammatical *forms*, speaking of medium and passive *meanings* instead, when their semantics is involved. The medium endings are, partly also depending on morphonology and inflection class: 1SG *-(o)mai, -(ó)mēn*; 2SG *-sai, -ei, -ē, -so, -ou*; 3SG *-(e)tai, -(e)to*; 1PL *-(ó)metha*; 2PL *-(e)sthe*; 3PL *-(o)ntai, -(o)nto*; 2DU *-(e)sthon*; 3DU *-(e)sthon, -(e)sthēn*. Passive forms, which only exist in the aorist and the future tense, are marked by the suffixes AOR *-thē-* and FUT *-thēse-, -thēso-*, followed by active endings in the aorist, and medium endings in the future. The easiest, expected pattern of readings for a transitive verb would be that the PRS.MED is ambiguous between the typical medium readings and a passive reading, whereas the AOR.PAS form can only have a passive meaning, and the AOR.MED form only a medium meaning. However, this distribution also depends on verb class, as Kaufmann (2004, 127-140) shows in detail. E.g., intransitive verbs allowing for a decausative reading of the PRS.MED often encode this reading by a passive form in the aorist; some AOR.PAS forms even retained active meaning, e.g. *dielékhthēn 'to discuss'*. Pre-Classical, e.g., Homeric, Greek (8th-7th c. B.C.) used passive forms (*-thē-* aorists) very rarely; they developed only until the period of Classical Greek (500-330 B.C.) (Kaufmann 2004, 140).

Traditional grammars commonly claim that the AOR.MED must have also covered the pas-

sive meaning at earlier stages, in line with a general course of development in Indo-European languages, namely, the replacement of an active/medium system by an active/passive system. Incidentally, the development in OCS can be seen in this connection (cf. section 5.5.2). As Kaufmann (2004, 141f; cf. also the references cited there) rightly criticises, this scenario leaves it as a mystery (i) why the present tense did not share the same fate, and (ii) why the new *-thē*-passives in the aorist should have developed from actives, rather than from media (as evidenced by their endings). She points out that most early alleged passive readings actually involve intransitive change-of-state verbs with an optional causer expression (*hypó*-PP or dative NP) – e.g., *αποθνήισκω (hypó) 'I die, am killed by'*. Medium forms with a passive-like interpretation first occur in the perfect, specifically, as descriptions of object properties. The perfect became a regular tense form in Classical Greek, establishing a dynamic reading besides the static result reading. This dynamic reading implies the presence of an implicit agent, leading to a truly passive reading of the medium. The passive reading then spread from the perfect into other tenses (Kaufmann 2004, 146). Further relevant changes obtained until and throughout the Koiné: In the aorist, a given verb came to be used only either in its medium or its passive form, which then had both medium and passive readings (Mussies 1971, cited after Kaufmann 2004, 161). In future tense, semantically empty medium forms corresponding to an active present form were replaced by actives, and *media tantum* by passives. The indirectly-reflexive reading of media was virtually lost. An important change is the frequent replacement of decausative media by actives, e.g. *dúō 'to sink (transitive or intransitive)'* < *dúō 'to sink (transitive)', dúomai 'to sink (intransitive)'*. When this change affected causative verbs, the medium became restricted to a passive reading.

Departing from her theoretical considerations, Kaufmann (2004, 159ff) considers various "strategies", which might have led to an integration of passive meaning into medium voice readings. According to "strategy A", the medium condition is weakened to extend to all cases of non-inclusion of the S-controller in the θ-grid. The medium would cease to be a marker of non-canonical control; as a consequence, the active-medium system could turn into an active-passive system, or the medium could become a general marker of intransitivisation (argument reduction). One-place *media tantum* should become reduced in frequency. "Strategy B" changes the medium condition to encode that the canonical controller is not linked to the highest available structural case, i.e., is not realised as the syntactic subject. This would lead to a systematic ambiguity between decausative and passive readings. If further developments served the goal of disambiguation, a passive marker could be introduced in order to mark blocking of the highest role from realisation, leading to a system of three voices; alternatively, medium voice could begin to mark only the blocking of the Agent role,

leading essentially to an active-passive system. "Strategy C" consists in a reinterpretation of S-control in order to cover also the Agent in a passive reading of medium voice (but not the Agent of actives). This strategy leaves the medium condition intact; however, it calls for evidence that Agent phrase in passive readings differs from the one in true passives.

Modern Greek faces a diglossia between the formal language (*katharévousa*) – essentially an artificial, partial revival of Ancient Greek –, and the vernacular (*dimotikí*). Interestingly, the passive reading of (earlier) medium[7] forms is generally considered as rare in the vernacular, and is preferrably replaced by topicalisation of the object with the verb in 3PL (Kaufmann 2004, 171) – a construction quite akin to Ru indefinite impersonals. Furthermore, overt Agent expressions in passive readings of the medium are often understood as instrumental or causative, and are restricted by the Agent's and the Patient's relative position on the animacy hierarchy. As Kaufmann shows, in Modern Greek only reference to a *specific* event requires that the Agent be present in the θ-grid as first participant; unspecific (e.g., generic) situations allow for S-control by an implicit agent. In sum, Modern Greek (vernacular) follows strategy C above and, in the long run, reduces passive readings again.

As a last step, Kaufmann applies her system to reflexive verbs in several Indo-European languages, including modern Ru. She cites the following possible readings of Ru *-sja* verbs: (i) directly reflexive, causative-reflexive, and (rarely and unproductively) indirectly reflexive; (ii) reciprocal; (iii) decausative (pseudo-passive and pseudo-reflexive); (iv) modal; (v) passive. There are *reflexiva tantum* among the verbs of emotion, positioning and movement, and among the inherently reciprocal verbs. Furthermore, reflexives have uses beyond the medial readings, as e.g. habitual (without object) or aspectual ones, and occur in e.g. in modal impersonals with an experiencer subject. The common denominator of all these readings seems to be *intransitivisation* (Kaufmann 2004, 217), leaving either only the Agent or only the Patient argument to be realised. Since both decausative and passive readings are (at least in imperfective aspect) fully productive, Kaufmann concludes that Modern Ru has followed strategy A in developing the former medium suffix into a general marker of non-inclusion of one of the structural arguments (not necessarily the subject!) into the θ-grid. Interestingly, Fehrmann et al. (2010) arrive at exactly the same characterisation of Ru *-sja*.

In a diachronic perspective, evidence from *Romance* languages points to the following path of development (Kaufmann 2004, 228ff, citing, among others, Selig 1998):

(60)   (i) directly reflexive > (ii) decausative pseudo-reflexive > (iii) decausative pseudo-passive > (iv) passive > (v) impersonal (transitive)

---

[7]Modern Greek has only two forms to express voice distincions, active and so-called passive; the passive is identical in form to the Classical medium and expresses the usual medium readings plus a passive reading.

Stage (ii) in (60) is connected to a reanalysis of verbs to become semantically intransitive: The predicate no longer expresses an activity, thus predicate control is excluded, but the only remaining argument is still an S-controller. A further reanalysis leads to (iii), which only contains a non-agent argument, the S-controller and predicate controller (causer) being removed from the thematic grid. *Reflexiva tantum* in Romance arose roughly at the time when reflexive decausatives became fully productive; this is expected under the view that reflexive morphology becomes a marker of medium voice. A modal reflexive construction comes with an implicit S-controller, contributed not by the lexical verb, but by a modalizer in the semantic representation. A passive obtains when the agent argument is in principle part of the SF, but existentially bound, so that the patient argument remains as the only one in the thematic grid to be projected into syntax. The passive reading developed after the modal reading. Thus, e.g. only two Latin examples of passive-like readings have been identified by Selig (1998, 37), both with modal semantics. While Spanish and Italian developed referential passive readings, French allows for reflexive passives only in generic statements (Selig 1998; Cinque 1988). The difference between passive and impersonal may be located in additional restrictions on the binding of the patient argument, an issue which Kaufmann notes only in passing. In which respects does this theory apply to Slavonic? At least two immediate predictions will be interesting for us in the remainder:

1. The path into passive meanings necessarily leads across decausative pseudo-passives.

2. If both decausative and passive meanings are fully productive, strategy A should apply; i.e., one-place *media tantum* should become reduced from the lexicon. – because medium marking is irrelevant if there is no argument to be excluded

5.4.2 Reflexivity and impersonality in modern Slavonic languages

In all Slavonic languages, reflexive markers are multifunctional: To name just some possibilities, they may signal reflexivity proper, as in the most prominent reading of Pol *myć się 'to wash (oneself)'*, or reciprocity, as in Cz *líbat se 'to kiss (each other)'*; they signal decausativity, as in Ru *dveri zakryvalis' 'the doors closed'* and passive or impersonal, as in Cz *říká se 'one says/it is said'*. The most basic grammatical subtypes of reflexive marking in Slavonic and other languages have been established in the discussion of Geniušienė (1987) and Kaufmann (2004) already (see above). We will stick to Kaufmann's terminology in the remainder. However, there are further uses of reflexive marking discussed in the literature on Slavonic languages (cf. the summary in Fehrmann et al. 2010, 205ff). These will be shortly discussed

in the present subsection, before we focus on the semantic analysis of Reflexive Passives and Reflexive Impersonals.

5.4.2.1 Further subtypes of reflexive marking

5.4.2.1.1 Antipassive (Deobjective)   In antipassives, a transitive verb is reflexivised with its object removed. The difference in meaning vis-à-vis an active clause mainly concerns the event structure of the verb. Rivero and Milojević-Sheppard (2003) analyse this construction as *Accusative Indefinite* and stress that the implicit object is restricted to arbitrary humans. In that sense, it forms the natural counterpart to Reflexive Impersonals, with REFL "replacing" the object instead of the subject.

(61)   a.   Deca   se   štipaju.   [Serbian]
            children-NOM   REFL   pinch-3PL
            'The children are pinching (somebody else).'

                                        (Progovac 2005, cit. from Fehrmann et al. 2010, 207)

       b.   Sobaka   kusaetsja.   [Ru]
            dog   bites-REFL
            'The dog bites (= is dangerous).'

       c.   Marek   się   bije.   [Pol]
            M.   REFL   beats
            'Marek fights (other people)'   (Rivero and Milojević-Sheppard 2003, 115 (40-a))

5.4.2.1.2 Prefixed resultatives   These can be formed from transitives (with the object removed) and intransitives, and involve a modification of the event structure such that the reaching of a termination point is implied. Since there are further, unpredictable changes in meaning, prefixed resultative are sometimes treated as seperate *reflexiva tantum*:

(62)   a.   Karel   se   najedl   (guláše).   [Cz]
            K.-NOM   REFL   eat-PT.M.SG   goulash-GEN
            'Karel ate goulash until he was full.'

       b.   Małgosia   wyspała   się.   [Pol]
            M.-NOM   sleep-PT.F.SG   REFL
            'Małgosia slept late.'

       c.   ubegat'sja, nabegat'sja [Ru] – 'to run oneself tired'   (Kaufmann 2004, 212)

5.4.2.1.3 Middles   Reflexive marking often brings about modal semantics, i.e., possibility or an evaluative flavour. It is instrumental to keep the constructions in which modal semantics is *obligatory* apart from those in which modality surfaces only optionally, sometimes due

to their generic reading. A typical middle (in Kaufmann's view) is (63), already cited above.

(63)   provoloka   gnetsja
         wire-NOM   bend-3SG
         'The wire bends / It is easy to bend the wire.'   (Kaufmann 2004, 205 (44-a))

Other authors stress the fact that a middle interpretation is strongly dependent on the presence of an evaluative adverbial. This is probably the standard view, taken e.g. also by Ackema and Schoorlemmer (2005).

5.4.2.2 *By*-phrases and the modification of thematic grids

On Fehrmann et al.'s (2010) account, the realisability of *by*-phrases is the most important feature contributing to the microtypology of reflexive marking in Slavonic. If a *by*-phrase may be realised, an open argument variable for the agent must still be available at the level of Semantic Form (SF). The referential identification of the agent with the syntactic adjunct is then achieved at the level of Conceptual Structure (CS).

The Slavonic languages fall into three groups With respect to the realisation of *by*-phrases in Reflexive Passive/Impersonal constructions. Group 1 (Ru and Belorussian) allows *by*-phrases with Reflexive Passives based on transitive verbs. At the same time, it does not form Reflexive Impersonals from intransitive verbs, or Reflexive Impersonals with an accusative object. Group 2 (Ukrainian, Upper Sorbian and Bulgarian) shares with Group 1 the realisability of *by*-phrases in Reflexive Passives, but also has Reflexive Impersonals from unergative verbs (without *by*-phrases). Ukrainian, furthermore, allows for Reflexive Accusative Impersonals with *by*-phrases. Group 3, comprising Pol, Cz, Slovak, BCS and Slovenian, possesses Reflexive Passives and Reflexive Impersonals from all verb classes, albeit generally without a *by*-phrase. Part of Group 3, namely Pol, Slovenian and spoken/dialectal BCS, also form Reflexive Accusative Impersonals (without a *by*-phrase).

5.4.2.2.1 Reflexive Passives   The possibility of *by*-phrases, i.e., the constitutive feature of Group 1, is attributed to a morpheme called *refl-1*. Its semantic effect in the case of Reflexive Passives is to remove the agent argument from the θ-grid and render the corresponding participant variable unbound. Syntactic derivation and semantic composition then proceed with this variable unaffected; it survives until CS, where it can be linked to the referent of a *by*-phrase (or otherwise, existentially bound by default, leading to an indefinite interpretation (*"someone"*)). *Refl-1* is thus *argument-blocking* (Fehrmann et al. 2010) in the sense that it makes an argument variable inert for syntactic realisation and semantics.

Fehrmann et al. (2010) extend the workings of *refl-1* to the blocking of the *patient* argument as well, which would leave a free variable for the patient to be identified or extentially bound at CS. Constructions for which this analysis should hold are the genuine reflexives and the Antipassive. I.e., genuine reflexives are analysed as involving an overt agent and a patient only conceptually identified with this agent; if an oblique argument is available (as in Antipassives), it can be linked to the free argument variable corresponding to the patient in an active clause.[8]

Refl-1 only applies to transitive predicates.[9] This makes the right predictions in two respects: First, only reflexives based on transitives allow for the realisation of a *by*-phrase, throughout the Slavonic languages. Second, in those languages which allow for *by*-phrase realisation, only transitives enter the Reflexive Passive, and Reflexive Impersonals are impossible (apart from the case of *acc/non-acc* verbs).

A second kind of reflexive morpheme, *refl-2*, covers the counterparts of Reflexive Passives in those languages in which *by*-phrases are illicit. Its effect is to either bind or restrict the highest argument variable, in both cases inducing an arbitrary human interpretation. In the subcase in which the variable is bound off by an operator, it becomes unavailable for identification at CS. In the second subcase, it only gets referentially restricted by a lambda operator, but still takes part in the syntactic derivation and semantic compsition. Cz Reflexive Passives (and Impersonals) would belong to the domain of the first subcase: The highest argument variable is bound off, and there is no syntactic subject. Pol Reflexive Passives (and Impersonals), on the other hand, are covered by the merely restricting subcase, such that they realise a (null) syntactic subject. Since this fills the argument slot, no *by*-phrase is possible. The assumption that *refl-2* only applies to the highest argument implies it cannot capture genuine reflexives and Antipassives in the same way as *refl-1* in languages of Group 1 does. Fehrmann et al. therefore propose a modified ("cut-in-half", so to speak) *refl-1* which only applies to the *lower* structural argument, for Group 3 languages.

5.4.2.2.2 Reflexive Impersonals   Reflexive Impersonals in group 3 languages (to which Pol and Cz belong) depend on the work of the reflexive morpheme *refl-2* (see above). It either quantifies over or restricts the highest available argument variable, leading to Reflexive Passives and Impersonals formed from all relevant verb classes – transitives, acc/non-acc verbs,

---

[8]An open issue seems to be why the interpretation of a genuine reflexive does not come out as involving a distinct, indefinite patient argument. Fehrmann et al. (2010) seem to assume that this is due to some defaul mechanisms at CS, or that there is indeed general ambiguity.

[9]There may be a problem lurking here in Fehrmann et al.'s (2010) analysis. *Refl-1* certainly applies to *predicates* (type *<s,<e,t>>*), but nothing seems to guarantee formally that these may be only transitive. Probably, one would have to assume that the resulting representation is senseless.

unergative and unaccusative intransitives. In the subcase of quantification, the argument variable is bound and thus inert for syntax, semantics, or CS, which leads to syntactically subjectless clauses with a semantic restriction on the highest argument, and no *by*-phrases. This holds for Cz. Syntax then ensures that in the case of transitives, the object of an active clause becomes the syntactic subject of the reflexivised verb. In the subcase of restriction, the affected variable remains available for syntactic instantiation by a null subject with human, arbitrary semantics, as in Pol. Syntactically, the sentence remains essentially active in diathesis, which facilitates Reflexive Accusative Impersonals.

Languages of Group 2 – Ukrainian, Upper Sorbian, and Bulgarian – also form Reflexive Impersonals from *intransitive* verbs. While *refl-1* takes care of the transitives here, it cannot be extended somehow to intransitives, because that would predict that *by*-phrases could be realised with intransitives in Group 2, contrary to fact. Instead, Fehrmann et al. (2010) propose a restricted version of *refl-2* to be at work in Group 2 languages. This *refl-2'* only pertains to the arguments of intransitives, which are operator-bound, and thus excluded from syntactic realisation. *Refl-2'* is more specific than *refl-2* in two respects: It cannot modify two arguments, and it cannot just restrict, but has to bind the argument to which it applies.

5.4.2.2.3 Implications   The approach offered by Fehrmann et al. (2010) makes no predictions for diachrony, but merely offers an adequate *logical* account of the current situation. An immediate implication of the distribution of properties over languages here is, however, that easy genealogical or areal generalisations fail blatantly: Group 1 comprises two East Slavonic languages, but not Ukrainian; group 2 is completely heterogeneous (Ukrainian, Bulgarian, Upper Sorbian), and group 3 collects the rest of West and South Slavonic, with one language from each family missing (because they belong to group 2). That part of group 3 which allows for Reflexive Accusative Impersonals (Pol, Slovenian, dialectal BCS) is also quite unrelated. So the main implication for a diachronic study is that changes must have occurred mainly for *language-internal* reasons and are often language-specific.

5.4.2.3 *Refl* **as an argument expression or a diathesis marker?**

Any account of reflexivity and argument structure has to face the issue which status should be attributed to the reflexive marker. Genuine reflexives seem to suggest an obvious parallel between a reflexive pronoun and an overt object:

(64)   a.   Valentýnka   myje   panenku.   [Cz]
           V.-NOM   wash-3SG   doll-ACC
           'Valentýnka washes her doll.'

b. Valentýnka   myje   sebe.   [Cz]
   V.-NOM   wash-3SG   SELF-ACC
   'Valentýnka washes.'

c. Valentýnka   se   myje.   [Cz]
   V.-NOM   REFL-ACC   wash-3SG
   'Valentýnka washes.'

But this superficial parallel may well be misleading, or hold only for a restricted set of subcases. The pattern in (64) breaks down, of course, for *reflexiva tantum*. (64-c) is also actually ambiguous between a genuinely reflexive reading and a Reflexive Passive (although the former is much preferred), which at least calls into question the *object* status of *se*.

The issue of whether *se* should be treated as an object, or at least as an argument, or rather as a diathesis marker (particle or postfix) on the verb, has had a long research tradition. Convincing evidence for the latter position was provided already by Havránek (1928). First, apart from the *reflexiva tantum*, there are reflexive verbs which do have a non-reflexive counterpart, but the reflexive marker, for various reasons, cannot be assumed to simply replace the direct object, cf. Pol *prosić kogo*-ACC / *prosić się kogo*-GEN *'to ask someone (for something)'* (Havránek 1928, 122) or Cz *radit se s někým 'to consult with someone'* vs. *radit někomu 'to advise someone'*. But even for those proper reflexives which stand in a fully transparent relation to a non-reflexive counterpart, the reflexive marker is not always equivalent to the "verb plus object" in terms of argument structure (Havránek 1928, 19, citing Ovsjaniko-Kulikovskij): Ru *ja tol'ko čto otlično vykupalsja 'I have just taken a gorgeous bath'* does not necessarily mean that I bathed myself; Ru *ona odevaetsja so vkusom 'she dresses with taste'* does not entail that she does the dressing on her own. (Cf. also Havránek (1928, 119-135) for ample evidence of asymmetries between reflexive and non-reflexive verbs related to each other.)

Second, there is syntactic evidence against analysing the reflexive marker as an object (in the cases where this could seem feasible at all) from agreement with secondary predicates. In older stages, the secondary predicate agreed in case with a non-reflexive pronominal object (65) as well as with the strong form of the reflexive pronoun (66):

(65) a. proč mě   buoh stvořil   slepa
         why me-ACC God   create-PT.M.SG blind-ACC.SG
         'why God created me as a blind man'

                                          (Gebauer 1970, 14th c., cit. Havránek 1928, 156)

      b. vydati sebe   Bohu   milého   dělníka
         turn-in SELF-ACC God-DAT dear-ACC worker-ACC
         'to turn oneself in to God as a dear worker'

                                          (BiblKral, 2.Tim. 2:15, cit. Havránek 1928, 157)

As far as short/clitic reflexives are concerned, this pattern is exceptional; Havránek provides some examples from Church Slavonic and Old Cz like (66):

(66) když ty mezi dvým zlým ležíš a nebezpečna se vidíš
when you between two-INS.PL evil-INS.PL lie-2SG and endangered REFL-ACC see-2SG
'if you are lying between two evils and feel endangered'

(BawEz, 1472, cit. Havránek 1928, 158)

The rule for verbs with a short/clitic reflexive marker is for the secondary predicate to agree with the *subject* as in

(67) řěka činí sě vodna
river-NOM make-3SG REFL watery-ACC
'The river becomes full of water'   (Hrad., 1360s, cit. Havránek 1928, 157)

(Cf. also Havránek (1928, 157ff) for further evidence from earlier and modern stages of Slavonic languages.) The finding carries over to historical Pol, cf. the following example from our 16th century subcorpus:

(68) [Król, co się w jego ekonomiach dzieje, nie wie, ani tym mu się zatrudniać przystoi:
– 'The king does not know what happens in his economies, nor is it his duty to be bothered with this.']

on się cały na usługi towarzystwa poświęcił
he REFL whole-NOM to services society-GEN devoted
'He devoted himself fully to serving the society.'   (StaPrze, PolDi)

In modern Cz, according to Havránek (1928, 160), accusative agreement is essentially possible only in language games, as a marker of *de se* (taking a perspective at the subject "from the outside").

Oliva (2001) takes up the arguments presented by Havránek and, likewise, concludes that *se*, albeit historically a pronoun of its own, has been grammaticalised as a reflexive marker, rendering its verb intransitive.

There are, however, also arguments *for* an independent status of reflexive markers: E.g., those Slavonic languages in which the reflexive marker is not a postfix on the verb, show the phenomenon of haplology: When two reflexive markers occur in the same clause, one of them is deleted. This holds in coordinations, but also e.g. in cases of clitic climbing in Cz:

(69) Olda se styděl představit novému řediteli.
O.-NOM REFL be-ashamed-PT.M.SG introduce new-DAT director-DAT
'O. was ashamed of introducing himself to the new director.'

Haplology is hard to explain if the reflexive marker is part of the verbal form. Hudousková (2010) proposes an sophisticated analysis of Cz reflexive marking which aims at deriving all of the empirical properties from the simple observation that *se* is a phonologically deficient element (a proper clitic), together with general assumptions about lexical insertion and semantic interpretation.

5.5 Indefinite Impersonals in the history of Slavonic languages

This section presents the main findings from the history of Ru, Pol and Cz in the realm of indefinite impersonals of the null pronominal (3SG/PL) and the reflexive subtype. The rather sparse remarks in the literature are discussed along with my own results from the available corpora (PROIEL, RRuDi, PolDi, DČNK and Vokabulář webový, and the parallel bible corpus).

5.5.1 Methodology

As already with referential null subjects in chapter 3, we will enquire the development of indefinite impersonals empirically from two sides: We will go the other direction and investigate relevant texts directly, focusing on the relative frequencies of the various means of expression of indefinite impersonals. Second, we will look at the development of the main *overt* exponents of the category – reflexive markers, in our case – and note changes in their spectrum of meaning and use.

The former approach will naturally also cover null expressions of the category – in the case at hand, null pronominal subjects with various agreement patterns. It is easy to see that none of the two perspectives is sufficient all by itself: Looking only for *overt* exponents of the category, we would have virtually no chance of capturing the null pronominal subjects (unless every single sentence with a finite verb were analysed). And looking only at running text, we would end up with sparse evidence: Although the reflexive marker is usually among the most frequent tokens of Slavonic languages, the relevant reading is comparatively rare.

As before, we will base our conclusions on stratified random samples rather than raw corpus data. This time, this will not only hold for the second approach, i.e., text samples from the corpus, but also for the approach based on overt exponents, simply because the amount of available evidence is so large: The Pol corpus contains about 43.000 and the Cz corpus about 60.000 instances of the reflexive marker. All of these were automatically extracted, together with a reasonable amount of context, and entered into a database, using a Python

programme. The database also contains basic metadata, such as year of origin, text title and author. Then, a random sample of 500 hits per half century and language was drawn within the database, and annotated in detail. The data is thus normalised for subcorpus size per century, facilitating direct comparisons between centuries. The investigation will allow us to answer the following questions:

1. How did the meaning and grammar of the reflexive marker in Ru, Pol and Cz develop diachronically?

2. How did various markers of indefinite impersonals evolve in relation to each other?

Parallel corpora provide a means of searching for a category with a zero marker in one language by querying an overt equivalent overt marker in another language, with the texts aligned to each other. If the parallel corpus embodies various diachronic stages rather than different languages, diacronic change should be detectable in this manner. In the present section, we will make use of this data source as well: For Cz and Pol, the parallel bible corpus fulfils this requirement. For Ru, the corpus of the *Puškinskij dom*, (http://lib.pushkinskijdom.ru), which contains Old Ru texts with aligned modern Ru translations, could be used, if only it were philologically more reliable. The main problem with these texts is that they are compiled from various manuscripts and editions in the interest of readability. On the other hand, editorial comments make the sources explicit, and can serve as orientation about the amount of unpredictable variation due to various sources. In our research project, we therefore filtered the texts from *Puškinskij dom* according to their reliability, leaving only those which were essentially due to one source or edition. These texts were compiled into a sentence-aligned parallel corpus using standard software (Tiedemann 2004), and then the modern Ru part was queried e.g. for passives, instrumental agent phrases, 3PL finite verbs in sentences without a nominative and compared to the realisation in the Old Ru part. The result is definitely not intended for statistical evaluation, because there are too many uncontrollable variables, such as choice of realisation in Modern Ru, tagging errors, alignment errors etc. Nevertheless, the material which we do find in this manner, may provide good positive evidence.

A classical, multilingual parallel corpus was used as evidence merely in the case of Greek and Old Church Slavonic. The central question concerning us here was whether Reflexive Passives, which can already be found in OCS, were a mere calque from Greek, as assumed by some scholars. Let us begin with this issue.

5.5.2 Old Church Slavonic[10]

The preserved OCS monuments contain numerous examples of Reflexive Passives based on transitive verbs with realized (70-a) or dropped (70-b) objects, cf.

(70) a. отъпоуштаѭтъ   сѧ   грѣси   твои.   [OCS]
          forgive-3PL   REFL sin-NOM.PL.M your-NOM.PL.M
          'Your sins are (being) forgiven.'   (Mt 9:2, Mar., PROIEL)

 b. тоѭ бо   мѣроѭ   еѭже   мѣрите   възмѣритъ сѧ   вамъ.
          that PTCL measure-INS.SG which-INS.SG measure-2PL measure-3SG REFL you-DAT
          'For with the same measure you measure it will be measured back to you.'

                                                                  (Lk 6:38, Mar., PROIEL)

Given the rarity of Reflexive Passives in Old Ru (cf. section 5.5.3) and Old Cz (cf. section 5.5.4), this fact comes as a surprise, and evokes speculations that OCS reflexive passives were simply a syntactic calque of medium and mediopassive forms in the Greek original texts, which later spread into the elevated, Church-Slavonically influenced registers of Old Ru. This view is implied by Janke (1960, 21): *"Man darf also annehmen, daß die russische Literatursprache auch ohne den Einfluß des Altbulgarischen und letztenendes des Griechischen früher oder später ihr reflexives Medium zu einem echten Passivum entwickelt hätte."* – 'One thus may assume that the Russian literary language would have developed its reflexive medium into a true passive sooner or later, even without the influence of Old Bulgarian *and, ultimately, Greek.'* [emphasis added, R.M.]; however, according to Janke (1960, 64), *"Durch die Übernahme der altbulgarischen Grammatik, die das Passivum ihrerseits nach dem Vorbild des weiter entwickelten Griechischen ausgebaut hatte, wurde die russische Literatursprache der Mühe enthoben, das Passivum aus den bereits vorhandenen Formen zu einem geschlossenen System auszubilden."* – 'Due to the borrowing of Old Bulgarian grammar, which had established the passive following the example of further developed Greek, the Russian literary language was saved the effort of forming a passive from the already existing forms.' Janke claims that Old Ru developed its (reflexive and periphrastic) passive under OCS and ultimately, Greek influence.

Classical Greek formally distinguishes passive and medium voice only in the aorist and future tense, and uses a so-called medio-passive form with passive interpretation in the other tenses. Changes before and throughout the Koiné period led to a further blurring of voice

---

[10]This section is a revised and extended version of section 3.1 of Meyer (2010). Neither version would have been possible without the professionality and generosity of my colleagues from the PROIEL project, Dag Haug and Hanne Eckhoff, Oslo (cf. Haug and Jøhndal 2008; Haug et al. 2009), who compiled and annotated a multlingual parallel corpus of early Indo-European languages, and granted me full access to the Greek and OCS XML data, which I most gratefully acknowledge.

distinctions in the aorist: a given verb came to be used only in either its medium or its passive form, which then had both medium and passive meanings.

The ambiguities might have acted as a bridge for OCS to develop a passive interpretation on its reflexive forms as well, assuming that a writer might have rendered Greek mediopassives and the aorist media with passive meaning, by OCS reflexives. A detailed comparison of OCS Reflexive and Periphrastic Passives vs. the aligned Greek original verb forms in Codex Marianus, however, shows the following:

1. OCS Reflexive Passives often occur as translations of Greek morphologically unam-
   biguous passives (not mediopassives), cf.

   (71) ι   оправьди   сѧ   прѣмѫдрость   отъ   чѧдъ   своихъ
         and justify-3SG REFL wisdom   by   children REFL-POSS-GEN.PL
         (Gr. ἐδικαιώθη)
         justify-AOR.PAS
         'And wisdom be justified by its children.'   (Lk 7:35)

   (72) въздастъ   бо   ти   сѧ   во   вьскрѣшение   праведъныхъ
         return-3SG PTCL you-DAT REFL in   resurrection   just-GEN.PL
         (Gr. ἀνταποδοθήσεται)
         return-FUT.PAS
         'For you will be repaid at the resurrection of the just.'   (Lk 14:14)

2. Koiné passive aorists with medium semantics – e.g. decausatives – are always marked by reflexives in OCS, never by participial passives. I.e., the conflation of forms in the original was not transferred to the translation. OCS remained true to the meaning, and did not blindly transfer form.

3. OCS already has reflexivisation strategies of its own, which can very well correspond to a non-passive form in Greek. Thus, the decausative in (73), and the "involuntary state" construction in (74) are both reflexive-marked in OCS, but correspond to Greek actives:

   (73) аште   ми   сѧ   ключитъ   съ   тобоѭ   оумьрѣти (Gr. δέῃ)
         if   me-DAT   REFL happen   with you-INS die   happen-PRS.SUBJ.ACT
         'if it will be my fate to die with you'   (Mt 26:35)

   (74) что   ти   сѧ   мьнитъ   симоне (Gr. δοκεῖ)
         what   you-DAT REFL thinks-3SG S.-VOC   think-PRS.ACT
         'What do you think, Simon?'   (Mt 17:25)

4. In a similar vein, OCS deponentia are generally quite independent of Greek (non-)
deponentia, cf.

(75) ı тъгда постѧтъ сѧ въ ты дьни. (Gr. νηστεύσουσιν)
and then fast-3PL REFL in those days fast-FUT.ACT
'and in those days, they will fast.'   (Mk 2:20)

Possibly, it would have to be more precise what is meant by borrowing of the Reflexive
Passive; the most plausible explication, on my view, would be that the semantic scope of
the reflexive marker was changed according to the Greek original, so that passives marked
by a medium form in Greek would be reflexive-marked in OCS. It has to be admitted that
this is the case: The vast majority of Greek mediopassives is indeed translated to OCS as
Reflexive Passives. But at the same time, the OCS system of reflexive marking was, at least
in important aspects, no mere copy of the Greek medium, as indicated by the observations in
1.-4. above. Another argument for the view that the OCS (Reflexive) passive system was to
a great extent autonomous, comes from Bräuer's (1952) study of Agent expressions in OCS
passives, covering Reflexive as well as Periphrastic Passives, both with present tense and
past tense participles, as well as impersonals. As he can show, the early translations of the
gospel consistently use *otъ 'from'*-P[repositional]P[hrase]s to (optionally) express an agent
in Reflexive Passives, but bare instrumental N[oun]P[hrase]s in combination with present
passive participles. Of course, one may doubt to what extent the instrumental NPs with
– usually resultative – present participles really encoded Agents. In periprastic passives
based on *past* participles, both *otъ*-PPs and NP-INS occur. But the point here is that the
Greek source model is the ὑπό-PP most of the time, so we are obviously facing a Slavonic
distributional rule. Later writings (Codex Suprasliensis and smaller documents) show more
variation and use both *otъ*-PPs and NP-INS synonymously in the same contexts (Večerka
1996, 227).

The view that there is no simple transfer between the expressions of passive from Greek
to OCS, is also advanced by Večerka (1996, 215ff), who states that the relation is asymmetric:
In the oldest documents (i.e., the gospel translations), the Greek synthetic passive is trans-
lated to OCS by Periphrastic or Reflexive Passive; however, the Greek periphrastic passive is
virtually never translated by a Reflexive Passive (Večerka 1996, 219). Furthermore, the Greek
passive is sometimes etranslated by a 3PL impersonal in OCS (Večerka 1996, 224).

Assuming that these arguments are on the right track, I conclude that OCS did not just
copy Greek when developing its Reflexive Passive, but rather followed a system of reflexive
meanings of its own (pace common heritage among the old Indo-European languages). Bor-

rowing from Greek does not solve problem why these Reflexive Passives are so rare in both Old Ru and early Old Cz. Let us now look at the evidence in OES, OCz, and OPol in more detail, before returning to this issue.

5.5.3 Old Ru

5.5.3.1 Reflexive and Periphrastic Passives

Janke (1960) claims that only those OES texts which were heavily influenced by Church Slavonic contain (agreeing or non-agreeing) SE-passives. Texts primarily based on the OES vernacular[11] allegedly only know single transitional cases ("Übergangspassiva" – 'transitional passives'), which should rather be interpreted as medio-passives – presumably reflecting the scope of usage of the respective Greek forms. In Kaufmann's (2004) terminology, these would be called decausatives (pseudo-passives). Pseudo-passives express spontaneous events without a situational controller; a causing situation may be inferrable, but is not part of the semantic representation of the predicate (Kaufmann 2004, 114f.). Janke (1960, 20ff.) seems to have a very similar concept in mind; he adduces examples like

(76)   хотѣлъ   быхъ   свободитися ѿт […] свѣта   сего
       want-PT.M.SG AUX-SBJ.1SG liberate-REFL from   world-GEN this-GEN
       'I would like to become free/be liberated from this world'   (IpatLet)

Consequently, Janke (1960, 21) views the reflexive passive in OES as a syntactic borrowing from OCS, although the East Slavonic vernacular was also developing its reflexive medium into a passive. But as we saw in the previous subsection, this only shifts the dilemma a step further back, because the evidence for a syntactic borrowing from Greek is all but conclusive.

Krys'ko (1995, 505f.) notes that Reflexive Passives are only attested for a relatively small number of verbs (types) in 12th-13th c. Old Ru: 47 out of 102 verb lexemes forming Reflexive Passives in his corpus occur mainly in (canonical) juridical texts, a stylistic markedness which is taken as evidence for their relative recency (in this specific meaning). Passive meaning could often only be inferred from the context, and agent phrases were expressed irregularly. Krys'ko thus subscribes to Janke's view that Reflexive Passives were actually borrowings from Old Church Slavonic. Note, however, that Krys'ko's latter two statements could just as well be made about modern Ru, at least for verbs which in principle allow for a decausative meaning: They have to be disambiguated between decausative and passive with the help of the context. The "irregularity of agent expressions" is likewise murky; the point

---

[11]Janke (1960) takes this to be the charters and those parts of the chronicles which have no especially literary or Church Slavonic context.

can only concern their realisation as such, since they almost exclusively occur as *otъ*-PPs (rather than instrumental NPs) in the data; see below and Zarickij (1961, 107ff.) for more details. This leaves the rather small productivity and stylistic restriction as criteria. From his thorough review of reflexive verbs, Krys'ko concludes that various reflexive meanings (including decausative pseudo-passives – *stradatel'no-vozvratnye glagoly*) had already been established, but a regular voice opposition had not yet developed.

Concerning participial passives, Janke (1960) again finds a strong impact of register on their distribution: Only texts influenced by Old Church Slavonic contain present passive participles or past participles in a process (rather than result) reading (on the issue of result vs. process readings cf. already Havránek 1928; 1937). Krys'ko (1995) this time takes issue with Janke's analysis: According to the distribution over genres in his corpus, participial passives cannot be, as a whole, borrowed from OCS. At the same time, however, they allow for variation in the realisation of an agentive adjunct – either by an *otъ*-PP or by a bare instrumental. This variation is again interpreted as a lack of fixation of the category of voice (Krys'ko 1995, 504). – As an independent strong argument against Janke's pure borrowing account in the case of Participial Passives, one may point to the broad availability of this construction in early OCz (Štícha 1985, 1988; cf. section), where they cannot possibly stem from OCS. Since the issue of register is so important for the case at hand, it would be instrumental to use a corpus of "proper" early OES texts, virtually uninfluenced by OCS. More or less reliable candidate documents in this respect include the Novgorod (and other) birchbark charters, charters on paper, and the chronicles. A full evaluation of the birchbark charters of Zaliznjak (2004) yielded only three relatively late instances of a Reflexive Passive from 14th c. onwards, cf.

(77) a. ваше бортико окралосѧ первы
         your   apiary-NOM.SG robbed-3SG.REFL first
         'Your apiary was robbed first.'   ([Gram Tver.5], 1300-1320)

      b. нинеце проноситсѧ ѿ кюрьяка
         now   propagate-3SG.REFL by K.-GEN.SG
         'now it is being propagated by Kjur'jak.'   ([Gram 538], 1380-1400)

The agent phrase in (77-b) is realised as an *otъ*-PP, as was the rule for OCS Reflexive Passives (see above). The chronicles contain more Reflexive Passive tokens; these are, however, again restricted to very few verbal lexemes (cf. Krys'ko's 1995) – *нарещи сѧ* 'be named', *(про)зъвати сѧ* 'be called', *ѧти сѧ* 'be conquered', *родити сѧ* 'be born', *крестити сѧ* 'be baptised' – with rare exceptions as in (78):

(78) да не изищетсѧ смьртъ ихъ ѿ кнѧзѧ . вашего .
     that not demand-3SG-REFL death-NOM.SG their by ruler your
     'May their death not be demanded by your ruler.'   (Lavr. let., RRuDi)

Agent phrases with Reflexive Passives are also always *отъ*-PPs in the chronicles, and they remain so at least until the early 16th c. Apparent exceptions, e.g. in the *Slovo i molenie Daniila Zatočnika* (13th c.), –

(79) Гусли бо страяются персты, а тело основается жилами.
     zither PTCL tune-3PL-REFL finger-INS.PL and body stabilize-3SG-REFL vein-INS.PL
     'For the zither is tuned by fingers, and the body is stabilized by veins.'   (RRuDi)

– are better analyzed as involving instruments rather than agents. Among the earliest reliable examples of instrumental agent phrases in Reflexive Passives are

(80) учение не от человѣка, ни за человѣка, но Исусом Христомъ
     teaching not by man-GEN.SG nor via man-ACC.SG but I.-INS Ch.-INS
     дасть ми сѧ
     give-3SG me-DAT REFL
     'my teaching is not given to me by man nor through a man, but by Jesus Christ'
                               (Reading menaia, 16th c.)

At the same time, *отъ*-PPs are still widespread:

(81) а от Бога не помилованъ, а от народа проклят
     but by God-GEN not beloved but by people-GEN cursed
     'but not loved by God, and cursed by the people'   (Domostroj, 16th c., RRuDi)

As noted above (Krys'ko 1995), early Reflexive Passives seem to be subject to lexical restrictions, which are no longer relevant in modern Ru. In the Domostroj (16th c.), we definitely find Reflexive Passives of many more verb lexemes than e.g. in Nestor's chronicle, cf.

(82) въ них же свѣщи пред святыми образы возжигаются, завѣсою
     in them PTCL candles before holy pictures light-3PL-REFL curtain-INS
     закрываются
     conceal-3PL-REFL
     '... in which candles are lit in front of holy pictures, (they) are concealed by a curtain'
                            (Domostroj, 16th c., RRuDi)

It is difficult to assess whether this effect is not a mere byproduct of the generally higher lexical variation in the later monuments.

(83) shows that the rather clear split on modern Ru which associates Reflexive Passives with the uses of the imperfective aspect and the Participial Passive with perfective, result readings (cf. Siewierska 1988), was not active in early Old Ru.

(83) идеже множьство стражющиихъ съпасени бывають
      where lot   suffering-GEN.PL saved   be-3PL
      'where a lot of suffering (people) are being saved'   (SkazBoGl, 13th c., RRuDi)

Considering putative Reflexive Impersonals, there is a clear distinction between the sub-type formed by transitive verbs with an elided object, and the one based on true intransitives. While examples of the former have been provided above for OCS (cf. (70-b), (72)), I could not find any instance of the latter, neither in OCS nor in Old Ru. If this is on the right track, Reflexive Impersonals based on intransitives never developed in the history of Old East Slavonic/Ru – and after all, they are ungrammatical in modern Ru, except as middles. The variety of Reflexive Impersonals with an obligatory dative experiencer is attested in OCS (cf. (74)), in Old Ru and in later stages:

(84) a. Аще кому   случится   въ враждѣ умрети
          if   someone-DAT happen-3SG-REFL in murder die-INF
          'if it will happen to someone to die by murder'   (KievPat, 13th c., RRuDi)

      b. И егда письмо изготовил,   занемоглось   мне   гораздо
          and when letter prepare-PT.M.SG become-ill-PT.N.SG me-DAT greatly
          'and when I had prepared the letter, I fell badly ill'   (Avv., 17th c., RRuDi)

5.5.3.2 3SG/PL indefinite impersonals

Non-reflexive indefinite impersonals were available from the early writings, albeit with more flexible agreement patterns, as Borkovskij (1968; 1978, 220f) shows. Especially in the *Pravda Russkaja* (1282), 3SG null subjects were widespread. However, apart form this minor agreement variation, the *neopredelenno-ličnye* 'indefinite-personal' sentences were available all through the history of Ru and showed no sign of diachronic change, Borkovskij claims.

Since we argued at length in chapter 3 that Ru diachronically developed from a canonical NSL to a partial NSL, the alleged stability in the domain of indefinite impersonals actually comes as a surprise. Following Holmberg's generalisations, we would expect that at the earlier, canonical NSL stage, Ru did *not* have 3SG/PL indefinite impersonals with a generic, speaker-inclusive reading – like present-day Cz and many other pro-drop languages. Instead, it should have employed mechanisms like Participial or Reflexive Passives to express this reading. Later on, when it had developed into a partial NSL, 3SG/PL indefinite impersonals

should have been extended to the generic, speaker-inclusive reading, as is characteristic of modern Ru. If, in fact, the situation remained as stable as Borkovskij (1978) claimed, while it did change with respect to referential pro-drop, this would constitute clear counterevidence against Holmberg (2010).

Reviewing Borkovskij's examples, however, we find that he does allude to some specific meanings of indefinite impersonals, which might be reinterpreted as semantic change. The main, gradual criterion applied by Borkovskij is to what extent the reading is distinct from a coreferential one and comes closer to a truly impersonal one. In this vein, he states that there are examples in which "the verb is in present tense and denotes an action which happens always, beyond temporal conditions, while the indefinite person, to which the predicate is directed, is everyone in general, without exception" (Borkovskij 1968, 83). The paraphrase contains typical features of a generic reading: lack of temporal boundedness and universal force. According to Borkovskij, such readings do not occur in the Pravda Russkaja (1282), but very rarely (altogether 9 times) in charters from the end of the 14th and from the 15th century. One doubtful instance may be found in the Smolensk charter from 1229 (Borkovskij 1968, 85). Otherwise the early occurrences in charters are circumscribed as "simultaneously pointing at the multiplicity and at the singularity of the subject" (*ukayzvajut odnovremenno i na množestvennost', i na ediničnost' dejstvujuščego lica* – ibid.), which is demonstrated by coreference with a singular pronoun. Judging by this wording and the respective examples, Borkovskij presumably means *existential* force, i.e., a non-specific indefinite reading, cf.

(85) оже бьють волного чл͞вка, платити за головоу ·Ӏ· гривенъ
     if hit-3PL free   man-ACC.SG pay for head-ACC.SG grivna-GEN.PL
     серебра
     silver
     'If someone hits a free man, he shall pay 10 silver grivnas per person.'

     (cit. Borkovskij 1968, 85)

In legends – specifically, the *Žitie Feodosija Pečerskogo* and legend of *Boris and Gleb* – the almost invariably occurring reading is that of an anonymous group of people; i.e., the non-generic, inferentially definite reading (Berger 1991) which may be rendered by English *they*:

(86) повѣдаша же ярославоу о   всемь семь
     told-3PL MP Ja.-DAT   about all   this
     'And Jaroslav was told all this.'   (SkazBoGl, Borkovskij 1968, 87)

A similar picture holds for the first Novgorod chronicle; the only example "possibly close to an impersonal" (*možno sblizit' s bezličnym*, ibid. 89) contains the verb *zvati* (*зов тъ*-3PL

'(they/one) call(s)'). In Nestor's chronicle, a few more examples fall into this category, most of them again containing *зов тъ*, but also its near-synonym *нарекоша*-3PL.AOR*'(they/one) call(ed)'* and others like

(87) Ярославъ [...] положи в стѣи Софьи. цр҃кви южe   созда самъ
      Ja.   put (them) in St.   Sofia church which-ACC.SG founded self
      [...] в неиже ѡбычныя пѣсни   Бꙋ҃   въздають.
            in which common songs-ACC.PL God-DAT offer-3PL
      'Jaroslav put (the books) in St. Sofia cathedral, which he had founded himself, in
      which one offers/they offer common chants to God.' (Lavr.let., Borkovskij 1968, 92)

Although this example is admittedly close to a generic inclusive reading, an existential, speaker-exclusive reading is possible as well; it cannot be ultimately determined which of the two is intended. This obviously changes in documents of the 16th century, especially in the *Domostroj*. Cf.

(88) и в осень капꙋстꙋ   солѧт и свеклꙋ   ставѧ и рѣпꙋ
      and in autumn cabbage-ACC.SG salt-3PL and beetroot-ACC.SG crop-3PL and turnip-ACC.SG
      и   морковъ   запасаю
      and carrot-ACC.SG stock-3PL
      'And in autumn, one salts cabbage and crops beetroot and stocks turnip and carrot'
                                              (Domostroj, 16th c., Borkovskij 1968, 94)

Borkovskij (1968, 94) states that in examples like these, no group of (anonymous) people can be denoted by the subject any more; the agent is "even more indefinite". But it is clear that an existential or speaker-exclusive interpretation is by no means intended here; by contrast, the advice given by the narrator definitely applies to him as well. Further examples of this type can be found in Ludolf's grammar (1696) and in a collection of 17th century proverbs (both cit. after Borkovskij 1968, 100):

(89) a. въ дуꙋраки поставлѧютъ человѣка которои богатство и честь
          in fools   put-3PL   man-ACC.SG who   wealth   and honour
          не ищетъ
          not search-3SG
          'A man who does not quest for wealth and honour is considered a fool.'
                                                                        (Lud. 1696)

      b. Даравому канюю в зубы не смотря
          gift   horse-DAT in teeth not look-3PL
          'Don't look a gift horse in the mouth.'   (Posl. 17th c.)

This is, of course, exactly the reading which is excluded in modern Cz and Pol (see above

section 5.2.2) and which we should deem possible in a partial null subject language, following Holmberg (2010).

While Borkovskij tries to link the occurence of one reading or the other exclusively to the type of text involved – e.g., the "group of people" being typical for narratives –, a diachronic development is easy to make out here: The readings which were available throughout history are those which are still present in the canonical null subject languages Cz and Pol today, i.e., mainly the existential non-inclusive reading and the group of humans / *"they*-reading".[12] A generic, speaker-inclusive reading first shows up in single instances in charters (15th c.), and more frequently in the 16th c. (according to the evidence cited in Borkovskij 1968). The earliest examples still allow for a different interpretation, but the mid 16th century occurrences from *Domostroj* are beyond doubt. Borkovskij (1978, 217-229) does not add any new categories or more detailed information on readings; the general statement is, as before, that the use of indefinite impersonals depends on genre, but was possible in principle since the earliest writings.

5.5.3.3 Corpus results

The text samples from RRuDi (each about 600 clauses in size, unless the given text is shorter; cf. chapter 3 for details) contain only a handful of indefinite impersonals. A graphic overview of their absolute frequency per year of origin of the respective sample is given in fig. 5.2.

The categories used here correspond to those common in the general linguistic literature on the topic. Most interesting in our context is the *generic inclusive* reading, which is not available in Cz or Pol. As is obvious from the graph, this reading first appears around the end of the 15th c. (*Afanasij Nikitin*) in our samples, and spreads in 16th-17th c., most strongly in the *Domostroj*, but also consistently in the other samples. Genre does not seem to be relevant here; examples stem from the Avvakum's *Žitie* as well as Kurbskij's letters or the *Vesti-kuranty*. The distribution rather fits Borkovskij's claim of a correlation with textual contents – the advice given in *Domostroj* or the philosophical ponderings of *Avvakum* apparently tend towards generic readings. Nevertheless, it is interesting to note that this reading is also diachronically restricted.

The most richly exemplified reading of 3PL impersonals is the inferentially-definite *"they"*-reading. It appears countless times in the *Vesti-kuranty*, when the sources of news are indicated. In the course of the discussion of Malamud (to appear a) above, we presented her

---

[12]In the terminology advanced by Padučeva (1975), Berger and Weiss (1987) and Berger (1991): the non-generic, newly introduced group reading and the non-generic, inferentially definite reading. Evidence of the generic, inferentially definite reading is hard to come by; (87) might be such a case.
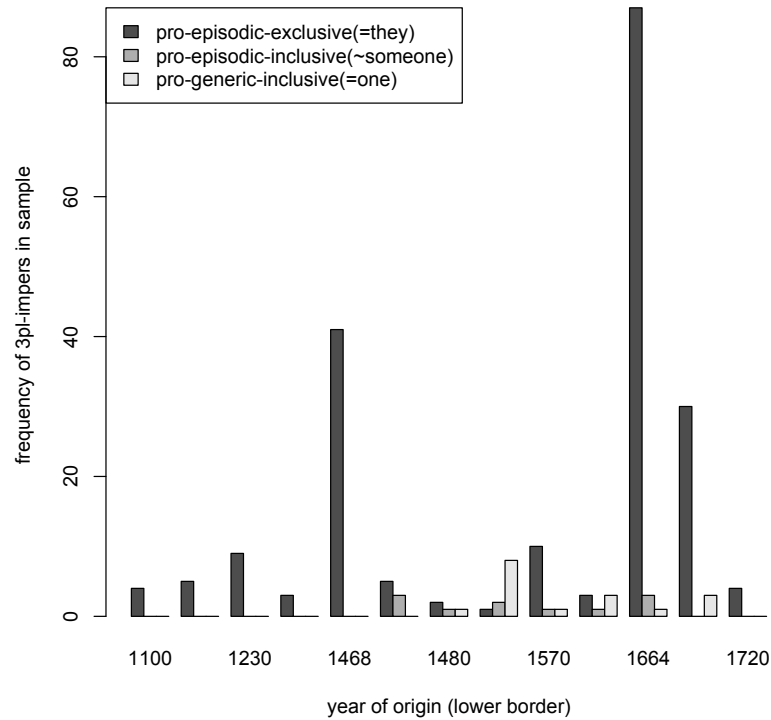
Figure 5.2: 3PL impersonals in the RRuDi samples

evidence from quantificational variability effects, pointing to the fact that 3PL impersonals behave like plural definites in modern Ru. Interestingly, as in Malamud's example (47), 3PL impersonals in the inferentially-definite reading could act as an antecedent and be taken up by a weak pronoun already in 17th c. Ru:

(90) Ис Полши пишѫт что они велми от приходу | тѹрскихъ
      from Poland   write-3PL that they very   of arrival-GEN.SG   Turkish
      людеи   богатца
      people-GEN.PL be-afraid-3PL
      'From Poland, they write to us, that they are very much afraid of Turkish people'

(Vesti 9, RRuDi)

5.5.3.4 Conclusion

Diachronic change in the realm of Reflexive and Participial Passives in Ru was very modest. Reflexive Passives apparently spread out and rose in productivity, and at a late stage

specialised towards imperfective aspect.[13] Agentive adjuncts took the form of *отъ*-PPs consistently since OCS until about the 16th c. Authoritative treatments in the literature claim that Reflexive Passives are borrowed from OCS (and ultimately, from Greek), against which we tried to argue. Reflexive Impersonals with a dative experiencer can be found early on. Kaufmann's (2004) theory would lead us to expect that, since the reflexive marker had already become a passive marker in OCS, it should not function as a marker of original medium meanings any more in Old Ru, contrary to fact. Kaufmann (2004, 217) argues that the reflexive marker in Modern Ru indicates that a structural argument corresponding to just any one of the predicate's thematic roles is not represented in the thematic grid (cf. also Fehrmann et al. 2010); reflexiva tantum and the original medium meanings then have to be treated as non-productive lexicalizations.

According to the evidence provided by Borkovskij (1968; 1978), and to the (limited) evidence from our corpus samples, it seems that 3PL indefinite impersonals changed their readings over time. While the inferentially-definite "they"-reading was available from the earliest documents, the generic-inclusive "one"-reading first appeared around the end of the 15th c. We found it to be, in principle, independent of genre, but possibly to some extent dependent on contents, as claimed by Borkovskij. In the present context, the diachronic change in readings is much more interesting than the latter point: The generic-inclusive reading of 3PL impersonals spread roughly at the same time as referential pro-drop (cf. chapter 3) began to decrease. This is, of course, in accordance with Holmberg's (2010) cross-linguistic generalisation concerning the readings of indefinite impersonals in partial vs. canonical pro-drop languages.

5.5.4 Old Cz

5.5.4.1 Previous research

According to Havránek (1937, 125-126), the Reflexive Passive (rather than the Participial Passive), represents the original Slavonic passive form. It is the most widespread expression of passive in the Slavonic languages (Havránek 1937, 3); on the other hand, the past passive participle has been used attributively and predicatively since the earliest sources (Havránek 1937, 71), both in its agreeing form and impersonally. Havránek (1937, 88) mentions Participial Accusative Impersonals already in the earliest Pol sources; there is no indication of this construction in OCS, but some single examples in later Church Slavonic. The temporal

---

[13]The details of this change demand a separate study which would also have to properly take into account the development of the aspect category.

meaning of this participle is resultative or perfect, but never that of an independent past tense. Havránek's analysis is (by intention, cf. p. 100f) very much influenced by a passage from Jan Hus' last work of writing, the *Postilla* (1413). In this passage, Hus explains his translation of the Latin passive *non iudicatur* as *nesúdie ho* 'not-judge-PRS.3PL *him*', pondering about various other possiblities: A Reflexive Passive *nesúdí sě* would tend towards a different meaning ('*táhne k jinému rozumu*'[14]); *nenie súzen* would correspond to the perfect passive *non est iudicatus*, and *nebude súzen* would be equivalent to the future passive *non erit iudicatus* (cit. after Havránek 1937, 101). Havránek concludes from this remark that the form *nenie súzen* had *not yet become* a present passive, but for Hus still carried a perfect/resultative meaning. He supports this claim by examples of translations of Latin perfect passives as participle plus present tense copula in Old Czech. On Havránek's view, this fact explains *proč se passivum v nejstarších textech tak často překládá jinak než opisným pasivem* 'why the passive in the oldest texts is so often translated in a different way than by a periphrastic passive'. The oldest translations of the New Testament allegedly use the Participial and the Reflexive Passive at an equal amount, whereas the later translations use the Participal passive at twice to four times as often as the Reflexive Passive. But the only actual example of an early Reflexive Passive adduced by Havránek (p. 101) involves the phrase *křtiechu sě ot něho* 'they baptise-AOR.3PL REFL *from him*'. This is a rather unclear case; it is possible that it should be translated as '*they had themselves baptised by him*', i.e. as a causative reflexive rather than a passive. – Havránek (1937, 103; 125f) claims that examples of Participial Passives denoting an ongoing action, rather than a result state, are very infrequent before the 16th century. Roughly from the 16th century, the perfect participle in combination with a secondary imperfective form of *byti* (e.g., *chválen bývá*) increases in frequency, as Havránek observes (1937, 125). He hypothesises that this type, such as the one with future copula (*bude chválen*) is supported by German influence, since it also occurs in Pol and Sorbian (before the loanword *wordować* 'to become' takes over in Sorbian).

The scenario developed by Havránek has been criticised repeatedly. Trávníček (1939) mainly addresses the direct comparison between Reflexive and Periphrastic Passives both by Havránek (1928; 1937) and by Gebauer (1929/ 2007). It is Trávníček's merit to have first worked out the point that Reflexive Passives are better described as a subtype of impersonals, which disallow *by*-phrases, and have an unspecific Agent. Therefore, they should not be treated as a true counterpart of the Periphrastic Passive, which allows the overt realisation of the Agent. On Trávníček's view, the Reflexive Passive (type *říká se*) developed directly from 3SG and/or 3PL pronominal null subjects (type *říká* and *říkají*), for which he finds Old Cz ex-

---

[14]Unfortunately, we are not being told *what* meaning.

amples as in (91), dialectal evidence before 1900, phraseologisms like Slovak *so statočnosťou najďalej ujde 'honesty is the best policy –* lit. *with honesty get-*3sg *farthest',* and some remains in modern Cz comparison constructions, as in (92).

(91) a. z   Babyloně putujíc   k svaté Kateřině na horu   Sinai
          from B.-GEN   travel-CVB.PRS to saint K.-DAT.SG on mountain-ACC S.
          musí   jíti
          must-3SG go
          'travelling from Babylon, one must go on mt. Sinai to St. Catherine's monastery'

          (mid 15th c., Mand., cit. Trávníček 1939)

      b. od   Sočavy k Dunaji počítají   mil   šedesát
          from S.   to D.   calculate-3PL mile-GEN.PL sixty
          'From the Soča to the Danube one calculates 60 miles.'

          (end 15th c., Kabátník, cit. Trávníček 1939)

(92) a. zapomnělo se   na to, jako kdyby kámen   do vody   hodil
          forget-PT.N.SG REFL on that as   if   stone-ACC in water-GEN throw-PT.M.SG
          'It was forgotten completely.'

      b. bolest   přestala, jako když utne
          pain-NOM stop-PT.F.SG as   if   cut-3SG
          'The pain stopped immediately.'   (Trávníček 1939, 16f)

He collects all of these subtypes under the cover term *manové věty* 'man *clauses'*, alluding to German *man 'one'.* Trávníček agrees with Havránek on the assumption that the Reflexive Passive is very old / original (*prastaré,* p. 15). He further points out – as earlier already Gebauer (1929/ 2007, 571) – that in Old Cz, agent phrases were sometimes compatible with the Reflexive Passive:

(93) Ježíš křtil   sě jest   od Jana
      J.   baptise-PT.M.SG REFL AUX-3SG from J.-GEN
      'Jesus was/had himself baptised by John.'   (1425, Koř., cit. Trávníček 1939)

With respect to Periphrastic Passives, Trávníček (1939, 19ff) essentially shares Havránek's view that the resultative or perfect meaning was historically primary to the actional use, which only developed in textually documented times. He argues that since the resultative meaning ascribes a property to the Agent, s/he naturally comes into the centre of attention more strongly. The Periphrastic Passive unites the resultative and actional meaning and retained this more central role of the Agent, which sets it apart from the Reflexive Passives/Impersonal. The Agent of the former is an individual one, the agent of the latter an indefinite one (p. 22). The change within the Periphrastic Passive described by Havránek is

related to the development of former perfect uses of all kinds of verbs into a grammaticalised aspect opposition, in which the perfective took over the resultative meanings, allowing the imperfectives to express (temporally simultaneous, not anterior) passives. Trávníček (1939, 21f) sets up some interesting analoga: A Reflexive Passive like *dům se staví* would have the active counterpart *dům stavějí lidé*; a Reflexive Impersonal like *zabil se*, however, would correspond to the active *zabilo ho (to)*, which also involves an indefinite Agent – the situation as a whole, an accident, an unspecified external force. In recent theories like Fehrmann et al. (2010) or Hudousková (2010), the latter case would be subsumed under decausatives; the precise relation to the subject of impersonals, however, still awaits an account. Since reflexive marking is multiply ambiguous in meaning, the Periphrastic Passive may sometimes be motivated by a need for clarity, because a Reflexive Passive might erroneously have been understood as a genuine reflexive. Interestingly however, some reflexive verbs – such as *pokládat se za 'to be considered as'* even favour the passive interpretation (p. 23).

Trávníček's (1939) contributes several remarkable insights: He establishes the major differences between Reflexive and Periphrastic Passives – the referential status of the Agent, and the realisability of *by*-phrases. The change within the Periphrastic Passive described by Havránek is more explicitly related to the development of verbal aspect. An interesting connection between decausatives and the "external force" Agents implicit in true impersonals is uncovered. And finally – in our context, most importantly – a new idea about the origin of Reflexive Passives/Impersonals is being put forward: Not only are they functional equivalents of earlier 3sg and (still existent) 3pl null pronominal indefinite impersonals, but their actual descendants. This claim ties in with very recent ideas in the theory of pro drop, cf. the discussion in section 5.3.2 above.

Trávníček (1961, 34ff) reiterates the earlier hypothesis from the perspective of null subjects with a "generalised" meaning. For Old Cz, these include a m.3sg (94-a), a 3pl (94-b), along with 1pl and 2sg null subjects.

(94) a. okolo toho města turkysóv najde mnoho
          around this-GEN city-GEN turquoise-GEN.PL find-3SG many
          'Around this city they/you/one find(s) a lot of turquoises.'

          (Mill., 14th c., cit. Trávníček 1961)

      b. mlýnov také v té zemi nemají
          mill-GEN.PL also in this-PRP land-PRP NEG-have-3PL
          'They also do not have mills in this country.'

          (Kabátník, end 15th c., cit. Trávníček 1961)

According to Trávníček (1961), the 3sg/pl cases (94-a,b) are frequent in Old Cz, but impos-

sible in modern Cz, where they have been replaced by reflexive constructions. The central idea, as before, is the genesis of Reflexive Passives/Impersonals out of pronominal null subjects.

Štícha (1985) takes issue with two of Havránek's main theses, namely:

1. The most widespread – and historically primary – form of passive in the Slavonic languages is the one expressed by reflexives.

2. Clearly actional, rather than resultative or statal uses of the Participial Passive are very infrequent before the 16th century.

Contrary to 1., Štícha (1985) observes that the relative amount of Reflexive Passives (compared to all passives) in his collection of 14th and 15th century Cz is in fact extremely low – between none at all and 7,5%; Štícha (1988) mentions 4% for the 14th and 7% for the 15th century, respectively.[15] What is more, its relative amount is on the rise diachronically. From the point of view of Old Cz, the first hypothesis is thus quite implausible. – Furthermore, Štícha (1985) presents some examples of actional readings of Periphrastic Passives already in 14th – 15th century sources, casting doubt also Havránek's second hypothesis. Unfortunately, it is often difficult to be sure about the resultative or actional meaning of the relevant examples. E.g., the paradigm example of a uniquely actional use discussed by Štícha (1985, 80) is the following:

(95)   [a jdúce na tu horu, tu stojí kapla,] – 'and going on that hill, there stands a chapel'

v   nížto leží veliký kamen, jenž byl   k hrobu božímu přivalen
in   which lies big   stone   which be-PT.3.SG to grave God-   rolled
'in which there is a big stone which was rolled in front of the grave of God.'

(Mand, 62, cit. Štícha 1985, 80)

According to Štícha (1985, 80), the reading can by no means be statal, because this would be equivalent to *"the utterance 'in which lies the big stone which lay in front of Christ's grave', which obviously makes no sense"* (– *výpověď 'v nížto leží ten veliký kámen, jenž ležel u hrobu božího', která zjevně nemá smysl*). The alleged contradiction is that the stone cannot be lying in front of the grave and in the church at the same time. But in fact, nothing at all precludes

---

[15]Actually, these numbers must be taken with a grain of salt, because they are based on a comparison with passive participle *forms* in all of their meanings (Štícha 1988, 24, fn.8: *nehledě na to, zda mají dějový, stavový či jakostní význam 'disregarding whether they have actional, statal, or qualitative meaning'*). Kotsmích (1869), who counted in only present indicative and past tense forms, finds slightly higher relative amounts of reflexives (cited after Štícha 1988). As far as the case against Havránek (1928, 1937) is concerned, the statal and qualitative uses of participial passives should not be included in the count, because Havránek's second hypothesis above already implies that these were frequent in the beginning.

that the result (= lying in front of the grave) was undone at a later time, before the stone made it into the chapel. Quite the contrary, the most plausible reading of (95) – as it appears to me – is exactly that of a pluperfect, i.e. the stone *had been rolled* in front of the grave: A stone which was at some time in the past being rolled towards the grave of Christ would not be relevant; what makes it interesting for believers is that it really lay in front of the grave. – A somewhat better case in point is Štícha's next example:

(96)　　… že   všichni vězňové,   jenž   byli   svázáni, byli   rozvázáni
　　　　　that all   prisoners-NOM.PL who-SG.M AUX-PT.PL enchained AUX-PT.PL unchained
　　　　a   jeho následovali
　　　　and him follow-PT.PL
　　　　'that all prisoners who had been enchained, were unchained and followed him'

　　　　　　　　　　　　　　　　　　　　　　　　(GestaB, 15, cit. Štícha 1985, 80)

Štícha reasons that a statal reading for *both* participal passives in (96) would imply that the prisoners were tied and untied at the same time. This is not true, because again, nothing enforces the temporal identity or overlap of the resulting states in the past – they might just as well follow each other.[16] What does suggest an actional reading *for the second passive* in (96), however, is its coordination with the actional *následovali* without any explicit indication of anteriority/posteriority. But again, this is an intuition which easily might be disappointed by some further factor in the context (maybe even *následovali* is a pluperfect?). Since no stronger criteria than intuitions like these are provided, I conclude that the case against Havránek's second hypothesis remains open. Moreover, Havránek never claimed that no actional readings whatsoever existed in 14th century Cz, only that they were far less frequent than statal ones.

Štícha (1988) once more scrutinises Havránek first claim above, this time on the basis of more quantitative and qualitative evidence. He points out that Gebauer (1929/2009, 512) had already taken the exactly opposite stance, namely that the spread of the Reflexive Passive was a later development, whereas the Participial Passive was more frequent in early texts. Trávníček (1939), however, had principally agreed with Havránek. The main function of

---

[16]Taking an example from German for comparison, it is completely unproblematic to have two pluperfects with contradictory result states next to each other:

(i)　Nachdem die Gefangenen, die   gefesselt worden waren, befreit worden waren, folgten
　　after   the prisoners who tied   been   were untied been   were followed
　　sie   ihm nach.
　　they him after
　　'After the prisoners who had been tied, had been untied, they followed him.'

the modern Cz Reflexive Passive – according to Štícha – is to signal a general, anonymous agent, who may be indefinite or definite, but unexpressed (cf. Trávníček 1939 above). The Participial Passive, on the other hand, presents the verbal action "from the perspective" of the object. This opposition in "perspectivity" developed around the turn of the 19th/20th century; earlier, both passive forms supported both perspectives. Unfortunately, no criteria for delimiting "perspective" are being provided. The corpus counts presented by Štícha (1988) suggest a slow, unsteady rise of relative frequency for Reflexive Passives (compared to all passive participle forms) since the 14th century: They were extremely infrequent before 1500 (4-7%), then leaped upwards during the 16th century (22,6%), and rose further to 37% in 18th and 27,2% in 19th cent., with great variation between individual texts. The expression of the agent by an overt *od*-PP (= *by*-phrase) was possible from the beginning until the 19th century. Judging from the presented evidence, Štícha (1988, 22) draws no distinction between Reflexive Passives and Reflexive Impersonals. We will see below that the spread of the construction across different verb classes represents an important diachronic change in the realm of Reflexive Passive/Impersonal. The central insight argued for by Štícha (1985; 1988) is that the Reflexive Passive should not be viewed as primary, because it is exceedingly rare in early texts; the evidence rather points to the Participial Passive being the original Slavonic passive (cf. Gebauer 1929/2009 for a similar view).

Agentive phrases – in the form of *od-* or *skrze*-PPs or bare instrumental NPs – can be attested since the 14th century. The earliest instances have been attributed to influence by the Latin original (*ā/ab*+ABL; Schäfer 1982, 167). Such an example would be (93) above, which virtually identically also appears in the *Bible Drážďanská* (2nd h. 14th c.; Štícha 1988, 25). Štícha's (1988) examples from several centuries show that the expression of the agent by an adjunct remained possible until the 19th century. Some examples – also from corpora – are given below; (97-c,d) suggests that a *by*-phrase could even accompany Reflexive Impersonals which were based on intransitives:

(97) a. Mnúť se již to nikterak déle nebude trpěti
          me-INS REFL-ACC already this no longer NEG-AUX.FUT bear-INF
          'I will not bear this any longer'   (end 14th., cit. Štícha 1988, 25)

      b. Tohoto Pavla Prvního schvaluje se od mnohých kronykářů k
          this-GEN P. first-GEN praise-3SG REFL-ACC by many chronicler-GEN.PL to
          chudým zvláštní přívětivost
          poor-DAT.PL particular cordiality
          'This Pavel I.'s particular cordiality towards the poor is praised by many chroniclers.'   (Glatouinus 1565, DČNK)

c. Ráno ... k modlení se od knězí volalo.
   morning   to praying-DAT REFL-ACC by priest call-PT.N.SG
   'In the morning, the priest called to prayer.'   (17th c., cit. Štícha 1988, 26)

d. Žádáme také, aby nyní, ... přísně se při tom od všech
   ask-1PL   also that-SBJ now   sternly REFL-ACC at this-PRP from all
   profesorů   pokračovalo
   professor-GEN.PL continue-PT.N.SG
   'We also ask that it was continued severely by all professors.'

(19th c., cit. Štícha 1988, 26)

According to the model by Fehrmann et al. (2010), the loss of *by*-phrases would represent an important diachronic change, the rise of a microtypological grouping with respect to reflexivisation among the Slavonic languages. Technically, this would involve the introduction of *refl-2.* One might reconstruct the development in a stepwise manner: *Refl-1* fully absorbs the λ-binder of the relevant argument variable, reducing the θ-grid of the corresonding active voice and leaving the variable unbound. *Refl-2* – in its first variant – only binds off the variable, without actually reducing the Θ-grid (= the Cz case at hand). A further, weakened version of *refl-2* could merely restrict the variable, rendering it available again for syntactic relations like binding and control, as in modern Pol. Metaphorically speaking, the highest argument would become more syntactically active in the process. The expectation would be that Pol had to run through a phase of being "like Cz" with respect to Reflexive Passives and Impersonals, before reaching the current situation.

A second prediction of this model is more problematic in light of examples like (97-c,d): If at some point, Reflexive Impersonals based on *intransitive* verbs indeed systematically allowed *by*-phrases, *refl-1* would not be sufficient to account for this period, because it can only apply to transitive verbs. Recall that *refl-1* changes two-place predicates of type *<e,<e,t>>* into one-place predicates with a free variable. Following this line of thought, it could not have been the modern *refl-1* that was at work in Old Cz, but an *argument-blocking* version *refl-2*". From the perspective of Fehrmann et al.'s (2010) theory, the most important point of Štícha (1988) thus clearly concerns his remarks about the development of *by*-phrases.

Before jumping to conclusions, however, we should first map which verb classes were involved in Reflexive Passives/Impersonals at all in the history of Cz. This issue has not been addressed before in any detail, neither by Havránek (1928) nor by Štícha (1985, 1988), who simply conflate the subtypes – cf. the impersonal (97-c) and examples based on intransitive *pokračovat 'to proceed'* and *ssednout 'to dismount'* amidst numerous examples formed from transitives in Štícha (1988, 26). Havránek (1928) keeps e.g. impersonals with an overt DAT experiencer apart logically, but does not make explicit any differences in diachronic devel-

opment between these and the Reflexive Passives (of which he gives only two examples, as Štícha 1988 rightly criticises). Schäfer (1982, 169) alludes to the issue in passing, when he concludes from Gebauer's (1929/2007) collection of examples that intransitives in deagentivised constructions have been in use only since the 15th century (*"Intransitive Verben werden in deagentisierten Konstruktionen erst seit dem 15. Jahrhundert verwendet."*).

5.5.4.2 Corpus evidence

We will first consider the development of Reflexive Passives/Impersonals with respect to verb classes. The earliest examples akin to Reflexive Passives in our Old Cz samples convey a permissive meaning, close to decausatives, and are restricted to a few verbal lexemes – *čisti 'read', psáti 'write'*:

(98)   Kapitola   padesátá šestá čte   sě: ...
         chapter   fifty-sixth-ɴᴏᴍ read-3sɢ ʀᴇꜰʟ
         'Chapter 56 reads / one can read in chapter 56: ...'   (PřIzJD, end 14th c., DČNK)

Impersonals with an obligatory dative experiencer are attested early on, as in OES above:

(99)   a.   jelikož   sě   jemu   jiesti chce
            how-much ʀᴇꜰʟ him-ᴅᴀᴛ eat   want-3sɢ
            'how much he wants to eat'   (ŽKrP, 2nd h. 14th c., DČNK)

       b.   [Ukradli jsme Janovi z Češnovic krávu, – 'We stole a cow from Jan of Č.,']

            a   z   té   mi   sě   dostalo   6   grošóv.
            and from this-ɢᴇɴ me-ᴅᴀᴛ ʀᴇꜰʟ-ᴀᴄᴄ get-ᴘʀᴛ.sɢ.ɴ 6   groš-ɢᴇɴ
            'and from (the sale of) this cow I made 6 grošs.'

                                                    (PoprRožmb, 1420, cit. Havránek 1928, 153)

The earliest reliable instances of Reflexive Passives, referring to actual rather than potential events, can be found already in documents of the 2nd half of the 14th century, albeit rarely. These are based on transitive predicates and belong to the agreeing subtype:

(100) a.   mast, kteráť   sě   jí   pomaže ve dne neb v noci
            cream which-ᴘᴛᴄʟ ʀᴇꜰʟ-ᴀᴄᴄ her-ᴅᴀᴛ apply-3sɢ in day or in night
            'a cream which is applied on her day or night'

                                                    (MastDrk, 2nd h. 14th c., VokWeb)

       b.   Toť jest krčemná příhoda, neb sě   pivo   pie,   a   ne
            this is tavern- event   since ʀᴇꜰʟ beer-ɴᴏᴍ drink-3sɢ and not
            voda.
            water-ɴᴏᴍ

'This is a tavern story, because one drinks beer and not water.'

<div align="right">(Podk. 1409, VokWeb)</div>

Reflexive passives from transitives with an alternatively sentential or PP object – Fehrmann et al.'s (2010) *acc/non-acc* verbs – also occur around the same time:

(101) po všem zajisté světě a ode všech se praví, že ...
over all surely world and by all REFL-ACC say-3SG that
'surely, it is said all over the world and by everybody that' (AlxB, 14th c., VokWeb)

Note in passing that (101) contains an overt *by*-phrase (*ode všech*). – Havránek (1928, 137) and Štícha (1988, 23) mention, among others, examples from the *Kniha rožmberská* (1st half 15th c.), a text which shows a relatively high amount of Reflexive Passives (11,6% of all passives, according to Štícha 1988). Štícha adduces more examples from the *Bible Drážďanská* (2nd half 14th c.), which cannot be due to borrowing, because the Latin of the Vulgata neither contains Reflexive nor Periphrastic Passives. Reflexive Impersonals based on *intransitive* verbs, however, remain rare exceptions even in the Kniha Rožmberská. They are a later development, which shows up regularly in our data only in mid 16th century:

(102) a. Pelagius začal, aby se památka za mrtvé dála. Lev Třetí, aby
P. began that REFL memory-NOM for dead-ACC give L. third that
se kadívalo, Innocentius, aby se (Pacem) líbalo.
REFL use-incense, I. that REFL peace kiss-3SG
'Pelagius introduced remembrance of the dead, Leo III. the use of incense, Innocentius the kiss of peace.' (Glatouinus, 1565, DČNK)

b. Když se tehdy v tom druhém place s koní ssedlo
when REFL-ACC then in this other place-PRP.SG from horses dismounted
[...]

'when one dismounted horses in that other place then'

<div align="right">(1599, cit. Štícha 1988, 26)</div>

c. jelo se velmi pěkně
drive-PRT.SG.N REFL-ACC very nicely
'one drove very nicely' (1644–1645, cit. Havránek 1928, 136)

d. Kdy se hnojiti má:
when REFL-ACC manure shall
'when it shall be manured' (Hosp 1705, DČNK)

Thus, the history of Cz shows, as did the current microtypology as well as the history of Ru, that the latter type has to be kept apart from Reflexive Passives based on transitive verbs.

236

A comparison of the frequencies of decausatives vs. Reflexive Passives from our database indicates that the former rise in relative frequency before the latter:
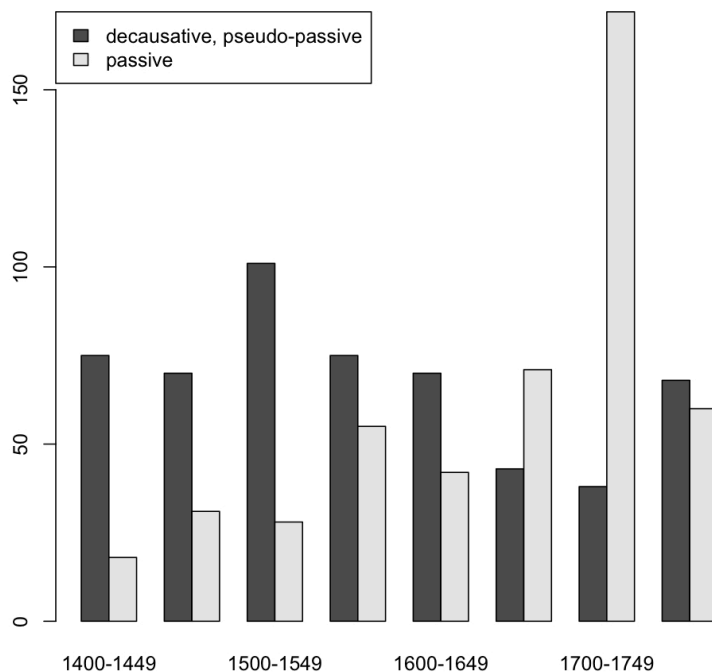
Figure 5.3: Cz: Decausatives vs. Reflexive Passives over time

This finding is in line with generalisations discussed by Kaufmann (2004) for Romance languages. It also jibes well with Haspelmath's (1990) developmental line from decausative to passive, at least for the period in which the Cz Reflexive Passive can still be analysed as a true passive.

5.5.4.3 3P null subjects and Trávníček's hypothesis

As has been repeatedly pointed out in the literature (cf. Trávníček 1956, 32ff), older stages of Pol and Cz made use of 3SG or 3PL null subject impersonals, which were later replaced by Reflexive Impersonals. The most direct illustration of this fact may be given by parallel versions of the same text from various time periods. In our small-scale corpus of various Bible translations, we find clear evidence:

(103)   a. Zdaliž sbírají z trní hrozny, aneb z bodláčí fíky?

              PTCL collect-3PL from thorn grapes or from thistle figs

              'Does one collect grapes from thorn bushes, or figs from thistles?'

                                            (BiblŠest, 1579-1593; BiblKral, 1613)

    b. Sklízejí se snad hrozny z trní a fíky z bodláčí?

              collect-3PL REFL-ACC perhaps grapes from thorn and figs from thistle

                                                (NovBiblKral, 2009)

However, influence by the original cannot be completely neglected here; 3PL indefinite impersonals even occur in the modern Cz part in meanings which we would not expect:

(104)   Také nerozsvěcují lampu, aby ji postavili pod nádobu, ale na

        also NEG-light-3PL lamp-ACC that-SBJ she-ACC put-PT.PL under bowl but on

        stojan

        pillar

        'One also does not light a candle in order to put it under a bowl, but on a pillar'

                                          (Mt 5:15, NovBiblKral, 2009)

(104) is clearly generic, without any delimitation of a potential set of referents; this reading is not usually expressed by 3PL indefinite impersonals in modern Cz. Thus, the majority of examples like (103) can serve as positive evidence of the fact that a change in the direction of replacing 3PL generic readings took place since the 17th century; but it cannot be guaranteed that the modern text always follows this development instead of using an archaism.

Note that among the examples given in the literature, the different readings are never systematically distingushed. Thus, e.g., Trávníček's (1956) above example (94-b) is non-generic, inferentially definite (akin to English non-generic *they*), and his (94-a) is ambiguous between this reading and a generic, inferentially definite one. Both could in principle occur with a null 3PL subject in modern Cz as well. But the situation is different with some of Trávníček's (1939, 16f) earlier evidence:

(105)   a. od toho města přijde k jednéj vlasti krásné

              from this-GEN city-GEN come-3SG to one country-DAT pretty-DAT

              'From this city, you/one get(s) to a nice country'   (14th c., Mill. 23a)

    b. púšče sú a lesové, skrze ňežto muóž svobodně jeti

              desert-NOM.PL be-3PL and forest-NOM.PL through which can-3SG freely go

              'there are deserts and woods, through which you/one can go freely'

                                          (14th c., Mill. 17)

    c. jablka velmi krásná, a když je rozřeže, tehda sú

              apple-NOM.PL very nice-NOM.PL and when them cut-open-3SG then be-3PL

        plna popela

        full-NOM.PL ash-GEN

'very nice apples, and when you/one cut(s) them, they are full of ashes'

(mid 15th c., Mand. 183a)

All three examples in (105) are of the "truly" generic type, without a delimiting adverbial. There is no reason to exclude the speaker from the set of potential subject referents. These examples are indeed in conflict with Holmberg's (2010) generalisations, assuming that Cz was a canonical pro-drop language at the time (mid 15th c.). There is not much doubt about the latter, cf. the referential *pro* in *tehda* (pro) *sú plna popela* in (105-c), or from the same text, the *Mandevilla*,

(106)   [...] kakžkoli jeden takově zchován bude,   že   nikdá   dále
              although one   thus   hidden   be-FUT.3SG that nowhere further
         nebýval,   však rád   slyší o   cizích   zemiech.   A   co   tu
         NEG-be-PT.M.SG still like-to hear-3SG about foreign country-PRP.PL and what here
         slyší   divného,   ješto jest   neznámé,   to   by   rád
         hear-3SG strange-GEN.SG which be-3SG unknown-N.SG that-ACC COND-3SG like-to
         spatřil.
         see-PT.M.SG
         'although someone will be hidden thus, that (he) was not anywhere far away, still
         (he) likes to hear about foreign countries. And what stunning (he) hears, which is
         unknown, that (he) would like to see.'   (mid 15th c., Mand., VokWeb)

If we assume that the Reflexive Passive rose in frequency and productivity since about the second half of the 16th c. (see section5.5.4.2 above), then Trávníček's hypothesis fits the chronology of changes quite well. The examples from the earliest Bible translations would either be archaic or mark the end of the era of generic 3PL null subjects in Cz.

5.5.5 Old Polish

5.5.5.1 Corpus evidence

As in Cz, the first examples of reflexives in a usage akin to passives mostly involve a modalized, permissive meaning, as in

(107)   Czcie się   tako we Księgach Sędskich: [...]
         reads REFL thus in books   judges
         'Thus it can be read/one may read in the book of judges: ...'

(KazŚw, m. 14th c., PolDi)

The first Reflexive Passives without this shade of potentiality or permissive causation are based on transitive verbs, and occur about a century later:

(108) prze[d] jimiż rzeczy dziedzinne i jine więcsz[e] wodzą się a
      before who things public and other bigger bring REFL and

      ro[z]prawiają
      discuss-3PL

      'before whom public things and other affairs are being brought and discussed'

<div align="right">(KodŚwięt, 2nd h. 15th c., PolDi)</div>

Again about a century later, we can identify the earliest Reflexive Impersonals, based on transitives with a PP complement, transitives used without an object (*odejmować* in (109-b)), or true intransitives (*chodzić* in (109-c)):

(109) a. [Lecz do tego trzeba pięknych zwyczajów, skromności przystojeństwa – 'How-
         ever, for this, nice habits and modesty of appearance are necessary']

         (od którego sie w żadnym postępku odstrzelać nie ma) …
         from which REFL in no case deviate not shall

         '(from which one shall not deviate in any case)'   (Dworzanin, 1566, PolDi)

    b. tak nieprzyjaciołom się odejmowało, tak sławy wielkiej u
       thus enemies-DAT REFL take-away-PST.SG.N thus fame big at

       postronnych dostało.
       uninvolved receive-PST.SG.N

       'thus the enemies were robbed and much fame was achieved among the unin-
       volved'   (Skarga, 1597, PolDi)

    c. kiedyby nie ta moja choroba, inaczej by się chodziło koło
       if-SBJ not that my illness otherwise SBJ REFL went-3SG.N around

       twej sprawy
       that problem

       'If there was not my illness, one/I would deal with that problem otherwise.'

<div align="right">(Listy, 1st h. 17th c., PolDi)</div>

Reflexive *Accusative* Impersonals are first attested in late 17th century and rise in frequency during the 18th cent.:

(110) A nawet i tych, co się ich tam i sam posyła,
      and even also those-ACC.PL that REFL them-ACC.PL thither and hither sends

      Tatarowie odzierają,…
      Tatars get-rid-of-3PL

      'And the Tatars even get rid of those, who are being sent here and there.'

<div align="right">(Sobieski, 1665-1683, PolDi)</div>

Pisarkowa (1984, 42) locates the rise of Reflexive Accusative Impersonals considerably later, namely at the *end* of the 18th c.; according to her, 16th-17th c. Pol knew only *agreeing* Reflexive Passives (i.e., not even Reflexive Impersonals based on intransitives). The remarks from historical grammars summarized by Szlifersztejnowa (1968, 166ff), however, support our above claim that Reflexive Impersonals dispersed during the 17th century and that Reflexive *Accusative* Impersonals came into use after those Reflexive Impersonals based on intransitive verbs had already been established.

More relevant than the respective earliest instances is the relative frequency of the above-mentioned subtypes over time. Our first evaluation is based on a random sample of 250 instances of the reflexive marker per 50 years, categorised for the reflexive subtype, agreement, presence or absence of a *by*-phrase, and presence or absence of an accusative object. 2328 instances altogether were analysed in-depth; the first sample from before 1450 only contained 128 hits. A certain dating problem arose with texts which could only be allotted to a time *span*, rather than one fixed year. Our solution was to use only texts whose time (period) of composition was fully contained in the relevant 50 years. Unfortunately, we only had *one* such text in the corpus for the first half of the 18th century (Chm); but since it was written in the 1740s, it may not even be characteristic of the early 18th century. So the results for the period from 1700-1749 have to be taken with a grain of salt.

A first quantitative result concerns the development of decausative vs. passive readings. As shown in diagram 5.4, the relative amount of decausative readings rises first, before the amount of passives increases. This is as expected according to Kaufmann's (2004) theory, and similar to Cz and to the findings for Romance languages (cf. Kaufmann 2004 for an overview).

This is not a trivial result, because there are other subtypes available which might easily spoil the picture, notably, the *Reflexiva tantum.* One may interpret the finding as indicative of an increasing grammaticalisation of reflexive marking, in the sense that the amount of lexically reflexive predicates decreases, and reflexivisation tends towards becoming a decausativisation and later, an impersonalisation marker.

Secondly, the diachronic distribution of the subtypes of *się*-Passives/Impersonals is of major interest. As illustrated in diagram 5.5, the three constructions rose in (token) frequency in the order *Reflexive Passives > Reflexive Impersonals > Reflexive Accusative Impersonals* in my samples (taking the Reflexive Accusative Impersonal in the 1st half of the 16th c. as an outlier):

There was a large number of ambiguous examples, which involved reflexivised transitive verbs in the N.SG and an argument with syncretic case marking or with no case marking at all. If the respective argument was taken to be NOM, the construction would count as a Reflexive
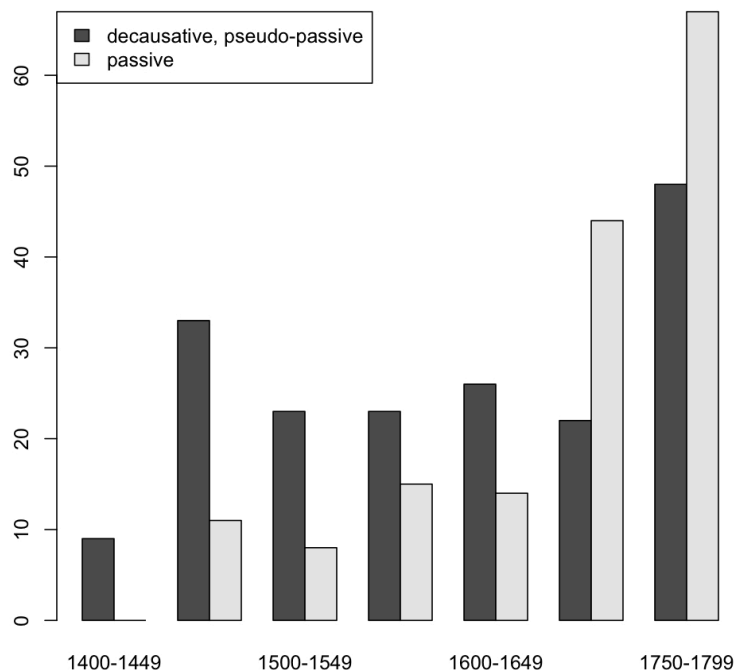
Figure 5.4: Decausatives and Reflexive Passives over time (Pol)

Passive, if it were ACC, we would have a Reflexive Accusative Impersonal. Unfortunately, examples like these can only be excluded from consideration.

5.5.5.2 *By*-**phrases**

As far as the possibility of *by*-phrases is concerned, Szlifersztejnowa (1968, 133) mentions instances of Reflexive Passives with agentive *przez*-PPs already from 15th-16th century. However, systematic investigation of all *się/przez* and *się/od* combinations in our corpus[17] yielded a different picture: Clearly agentive *by*-phrases are exceedingly rare with Reflexive Passives. This is indeed striking when we compare it to Participial Passives. Second, virtually all of the relevant examples would also allow for an interpretation of the referent of the *by*-phrase as an instrument, source or vehicle. It is often difficult to discern these, cf. examples like the following:

---

[17]We evaluated in detail all instances in which *się* and *przez(e)* or *od(e)* were located at a reasonable distance of maximally 6 words from each other, with no intervening sentence boundaries, marked up by JAPE rules and queried in GATE (cf. chapter 1). Recall that the base corpus was 4 Mio. tokens large, with about 43.000 occurrences of *się*.
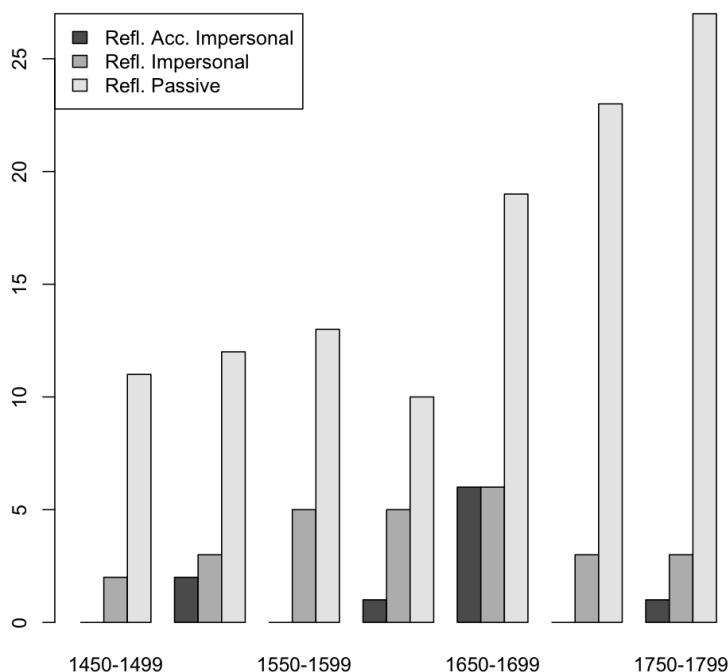
Figure 5.5: Subtypes of Reflexive Passives/Impersonals over time (Pol)

(111)   a.   [*Tę radę i tę tajemnicę chciał zjawić napirwej Gabryjełowi archanjołowi, –* 'This advice and secret, he first wanted to disclose to the archangel Gabriel,']

przez niegoż się moc boża rozumie
by/via whom refl power godly understand-3sg
'by/via whom God's power is being understood' (RozmPrz, end 15th c., PolDi)

b.   A pieniężnej piechoty nie używać, bo się skarb niszczy barzo
and squad not use-INF because REFL treasure destroy very
przez nię
by it
'and not make use of the squad, because the treasure is being very much destroyed by it' (Starowolski, 1618, PolDi)

c.   [Ta sztuka niedługo trwała, bo – 'This feat did not last long, because']

przez oszukanych prędko się po Warszawie rozgłosiła.
by/via deceived quickly REFL over Warsaw propagate-PT.F.SG
'it was quickly propagated over Warsaw by/via the deceived'

(Kit, 18th c., PolDi)

243

Many predicates are systematically ambiguous between a Reflexive Passive and an analysis as unaccusatives with an (optional) indication of a source:

(112)   a.   I   sam dobry uczynek, gdy się przez nas komu,   a
          and itself good   deed   when REFL by   us   someone-DAT and

          zwłaszcza wszytkim, dobrze czyni, ma swoje serdeczną zapłatę i
          moreover   all-DAT   well   does   has its   heartily   meed   and

          pociechę   wnętrzną wielką.
          consolation-ACC inner-ACC   great-ACC

          'And the good deed itself, when by/via? us good things happen to someone, and particularly, to everybody, has its inherent meed of the heart and great consolation.'   (Skarga, 1597, PolDi)

       b.   [...,] iż   się   wielkie złodziejstwa dzieją   od   Katalianow na
              that REFL great   crimes-NOM.PL happen-3PL by/through K.-GEN.PL   at

          morzu Czarnym i   Białym.
          Sea   Black   and White

          '[...,] that great crimes happen by/through the Katalians on the Black and the White Sea.'   (JanPam, end 15th c., PolDi)

The single most frequent verb which *might be* analysed as involving an agentive *przez*-phrase, is *wyrażać 'to express'* (20 examples, all from Chm, 1755/56):

(113)   Memoria,   wyraża się przez osobę o dwóch twarzach
        remembrance express REFL by   person of two   faces

        'Remembrance is expressed by a person with two faces.'   (Chm, 1745, PolDi)

However, since *active wyrażać* already selects for a *przez*-PP adjunct, and several of the aforementioned passive-like instances clearly contain an unintentional, inanimate, instrument-like *przez*-PP, I conclude that these examples should be disregarded. If we thus strictly include only *by*-phrases which contain unambiguously agentive highest arguments, we are left with a handful of examples (6 x *przez,* 9 x *od*):

(114)   a.   Zbierał się sejm, ale już   było   wiadomo, że   się
          gathered REFL Sejm   but already be-PT.N.SG known   that REFL

          gotował   przez niego okropny, a   w dziejach naszych
          prepare-PT.M.SG by   it   disgusting and in history   our

          niesłychany zamach na ojczyznę.
          unheard-of   assault   at   home country

          'The Sejm gathered, but it was already known that a disgusting assault against our home country, unheard of in our history, was being prepared by it.'

                                        (Sop, 1839-41, PolDi)

b. [...] których od   dawna   nie widział,   bo   od sejmu
whom   from long-ago not see-PT.M.SG because by Sejm-GEN
podziałowego zamknął   się był   w Morafie
division-GEN   lock-up-PT.M.SG REFL be-PT.M.SG in M.-PRP
'[...] whom he had not seem for a long time, because he had been imprisoned
by the Sejm division in Morafa'   (Sop, 1839-41, PolDi)

Even worse, those few reliable examples cluster in subgroups:

- We found 7 times causative *dać się*, (3 x *przez*, 4 x *od*). Its compatibility with *by*-phrases
  clearly distinguishes it from the common Reflexive Passives; *dać się* obviously rather
  behaves like a "true" Periphrastic Passive.

(115) a. Nie damy   się owładać przez pięciu posłów, konfederacji
not give-1PL REFL rule   by   five   delegates confederation-GEN
chcemy i   Ponińskiego za jej marszałka!
want-1PL and P.-GEN   as its marshall
'We will not let ourselves be ruled by five delegates, we want a confederation
and Poniński as its marshal!'   (Sop, 1839-41, PolDi)

b. [Z Londynu 14 ianuarii 1661 – 'From London, January 14, 1661']

Dała   się królowa matka przepraszać od   syna   książęcia de
give-PT.F.SG REFL king-   mother excuse   by   son-GEN duke-GEN   of
Iork z strony jego ożenienia   [...]
York regarding his   wedding-GEN
'The king's mother let herself be excused by her son, the Duke of York, con-
cerning his wedding ...'   (Mer, 1661, PolDi)

- 3 of the remaining 5 instances of agentive *od* are based on the verb *nagradzać 'reward'*.
  They might be "explained away" as true impersonals with a dative experiencer:

(116) Tak mi się moja ludzkość od   ciebie nadgradza.
thus me REFL my   humanity from you   reward-3SG
'This is how my humanity is rewarded by you.'   (Op, 1650, PolDi)

This leaves us with 6 examples altogether, one or two of them doubtful, stemming from
five different texts – a neglectible amount, especially in comparison with the Participial Pas-
sives and Impersonals, and with the situation in closely related Cz. I conclude that to the
extent that *by*-phrases are indicative of true Passives, they do not provide enough evidence
to support a slow drift from Passive to Impersonal in the history of Pol.

5.5.5.3 Pro-*arb* and diathesis

Turning now to our main line of investigation, namely, the distribution of pronominal (3SG/PL) null *arbs* vs. Reflexive Passives/Impersonals, a first thing to note is that the "*they*-reading" (excluding the speaker) is well-documented already at the beginning of the 16th century. The following example contains a whole sequence of such pro-*arbs*:

(117)   [Inszych czasow, procz oblężenia, każdy s służby swej żyje, – 'In other times, without siege, everybody lives on his service,']

a   służbę   im   na każdą   ćwierć roku   w pełni   od   dworu
and   service-ACC.SG they-DAT to every-ACC quarter year-GEN in full-PRP from court-GEN
przysyłają bez   chyby;   szatę   im   też dawają jednę   w rok
send-3PL   without failure-GEN dress-ACC they-DAT also give-3PL one-ACC in year-PRP
[...]

'and they are being sent service for each quarter of a year from the court, without failure; they are also given one dress per year [...]'   (JanPam, end 15th c., PolDi)

Examples with a *generic, speaker-inclusive* reading are rare and, of course, hard to identify with any certainty. I found the clearest instance in early Bible translations, where the original additionally helps with interpretation:

(118)   ognie[m] jeś   nas pokuszał,   jako pokuszają śrzebro.
        fire-INS   AUX-2SG us test-PT.M.SG like test-3PL   silver
        'With fire you tested us, like one tests silver.'   (PsFl, 1st h. 15th c., PolDi)

But (118) is not absolutely reliable; a generic reading with subjects in an inferentially definite set might be conceivable, and accordingly, the example may well be translated using *they* (rather than *one*) in English.

It is often the case that a 3PL pronominal pro-*arb* in an early translation corresponds to a Reflexive Impersonal in a later version.

(119)   [Mt 7.16]

a. izali zbierają z   ciernia grona winne, albo z   ostu figi?
    PTCL collect-3PL from thorn   grape wine- or   from thistle figs
    'Does one collect wine grapes from thorn bushes, or figs from thistles?'

                                                                    (BiblGd, 1632)

b. Czyż zbiera   się   winogrona z   ostu, a figi z ciernia?
    PTCL collect-3SG REFL wine grapes from thorn
                                                                    (BiblPoz, 1991)

(120)   [Mt 10.29]

     a.   Izali   dwu   wróblików   za   pieniądz   nie   przedają?
          PTCL   two-ACC   sparrow-ACC.PL   for   coin-ACC   not   sell-3PL

                                                             (BiblBrz, 1563)

     b.   Izali   dwóch   wróbelków   za   pieniążek   nie   sprzedają   [...]
          PTCL   two-ACC   sparrow-ACC.PL   for   coin-ACC   not   sell-3PL

                                                              (BiblGd, 1632)

     c.   Czyż   nie   sprzedaje   się   dwóch   wróbli   za   jednego   asa,   [...]
          PTCL   two-ACC   sparrow-ACC.PL   for   coin-ACC   not   sell-3PL

                                                              (BiblPoz, 1991)

One may rightly argue that Latin influence on the earliest translations was stronger than it is today – and the Vulgata, as a rule, uses a ACT.3PL in the respective places (*colligunt* 'they collect'; *veneunt* 'they sell'). But as in the Cz case above, it seems noteworthy that more recent translators found it necessary to introduce a reflexive construction. The reading of the examples in (120) is clearly generic, potentially including the speaker and hearer; an English translation with *one* would be standard. Recall that this is the reading which is missing from 3PL indefinite impersonals in Pol today.

5.5.5.4 Generalisations

The observations on Pol support the following generalisations:

1. Decausative pseudo-passives > Reflexive Passives > Reflexive Impersonals > Reflexive Accusative Impersonals

2. *By*-phrases were very rare from the beginning, if at all possible. It seems that the "Reflexive Passive" was Impersonal (only based on *transitive* verbs) from the outset.

3. Reflexive Passives replaced ACT.3PL pro-*arb* as the major means of expressing indefinite impersonals. It also became the *only* means of expression supporting a generic reading. This implies a change in meaning for the pro-arb, which became only coreferential (or a *specific* indefinite).

5.6 Summary

The present chapter covered a large array of facts in the synchrony and diachrony of indefinite impersonals and the closely related reflexive-marked passives and impersonals. Starting

from the basic observation that modern Ru, Pol and Cz differ on the readings assigned to indefinite impersonals, an overview of semantic and syntactic properties of both groups (null pronominal and reflexive) was given. General linguistic theories about indefinite impersonals have established a similar distribution of readings, together with a similar behaviour with respect to referential pro-drop, as we find in Ru, Pol and Cz today. Current analyses of reflexive-marked diathesis in Slavonic and other languages allow for some predictions for diachrony, mainly regarding the development of passive and impersonal meanings out of decausatives. In a corpus study of reflexive markers, their readings and syntax, in OCS, Old Ru, Old Pol and Old Cz, we found confirmation of the idea that passive meanings develop from decausative ones: The latter rose in relative frequency before the former began to increase. Concerning the earliest phase of development, OCS contains ample evidence of Reflexive Passives, whereas Old Ru and Old Cz virtually lack them in the beginning. While this is stunning in itself, it also shows that the overall lack of *any* passive structures in Old Ru does not imply that they were later imported from OCS, and, ultimately, from Greek. As I tried to argue, the borrowing thesis from Greek into OCS itself lacks empirical and conceptual support. Furthermore, if these structures were rare also in early Old Cz, but developed "organically" later, then the borrowing thesis loses even more plausibility.

More interestingly, the uses of reflexive-marked structures seem to have differed from the beginning: In Ru, Reflexive Passives are attested early on, whereas in Cz and Pol, these develop only in 15th-16h c. In Old as well as Modern Ru, they can contain an agentive adjunct (with a minor diachronic change in form), whereas in Cz, this was possible only until the 19th c., and in Pol, they have always been rare (and impossible today).

The central result of the present chapter, however, is to be seen in the empirical support in the realm of diachrony, for the hypothesis of a close relation between pro-drop behaviour and the availability of generic inclusive readings with 3P null pronominal subjects (Holmberg 2010) on the one hand, and with the use of Reflexive Passives in canonical pro-drop languages on the other (basically already Trávníček 1939; 1951; 1961; in connection with pro-drop Fassi Fehri 2009).

6 Conclusion and Outlook

The present thesis consists of a short technical/methodological chapter and three long linguistic chapters. This does not quite accurately mirror the effort that went into each of the parts, which was actually roughly equal for the technical and for the linguistic one. The groundwork of coding and "engineering", before being able to do anything linguistic, was necessary, because Slavonic diachronic corpora are still a rarity and lag far behind the resources available for Germanic or Romance languages. If the present work was able to demonstrate the usefulness of even imperfect diachronic corpora and the urgent need for more of them, then it has reached its first goal.

Linguistically, the following generalisations could be established, mainly on the basis of random samples drawn from the corpora:

1. Ru changed from a canonical pro-drop language to a partial pro-drop language in the course of the 16th/17th century. This change happened in parallel with the impoverishment of the preterite tenses, i.e., the loss of the 1/2P auxiliary and of the Aorist. However, 3RD person null subjects remained frequent for a longer time after the changes in the past tense system had taken place. This may be due to a syntactically informative zero auxiliary, which continued to be assumed by native speakers. Weak subject pronouns rose in relative frequency during the 17th c., indicating the change towards partial pro-drop. – It was crucial to compare null subjects to those subjects which *could* have been dropped, which had to be decided on the basis of referential, information-structural and syntactic criteria.

2. Cz and Pol remained canonical pro-drop languages throughout the 13th/14th – 18th c. There were minor changes of the licensing conditions for pro-drop over time, but without any clear direction.

3. Expletives in Cz and Middle Pol have been argued not to originate from an interjection or particle which was integrated into the clause and began to agree, as had been argued by Cz grammarians. The counterarguments were conceptual as well as empirical. The assumed change would be an unexpected degrammaticalisation, and it is not at all evident from the available historical data. Instead, it was proposed to view the attributive demonstrative as the historical source of the agreeing Cz expletive. An important role in the argument was played by Pol expletives, which were once common, but then got lost (before the 17th c.).

4. The chapter on indefinite impersonals was centered around two recently proposed generalisations concerning the possible readings for Reflexive Impersonals in canonical

pro-drop languages, and the difference between partial and canonical pro-drop languages with respect to readings of 3PL impersonals. Both generalisations were shown to apply also to Ru, Pol, and Cz. In the diachronic domain, it could be demonstrated that the crucial generic speaker-inclusive readings rose in frequency in Ru, as it evolved into a partial pro-drop language. No such changes took place in Cz or Pol. What did change in the latter languages, was the syntax of reflexive constructions. I try to relate the direction of this change to a recent theory of medium readings in typologically diverse languages.

It is obvious that these points leave a lot more to be explored and considered. I would like to highlight two major points. First, it would be very desirable to apply proper statistic methods to the diachronic data in full. The technique of logistic regression (Kroch 2000, Baayen 2008) was applied in chapter 3 and proved successful, but it could certainly also be useful for the analysis of change in the meaning of indefinite impersonals and possibly other phenomena. Unfortunately, given the amount of work on many tasks connected with data preparation and evaluation in the present thesis, some of these possibilities had to be left for the future.

Second, the extension of the time frame at which we looked was not quite sufficient, for some phenomena. Given that the Pol change towards more weak subject pronouns must have occurred during the 18th c., further detailed research within a time frame between the 18th and 19th c. might be rewarding. In this period, we would also expect to see another important change which could not be addressed in this work, namely, the rise of the restriction of Pol Reflexive Passives to a medium reading. I conclude with the hope that the present thesis at least methodologically paved the way for such research in the future.

List of Figures

List of Tables

Bibliography

Abraham, W., 1991. Null Subjects: From Gothic, Old High German and Middle High German to Modern German. From Pro-Drop to Semi-Pro-Drop. Groninger Arbeiten zur Generativen Linguistik 34, 1–28.

Ackema, P., Schoorlemmer, M., 2005. Middles. In: Everaert, M., van Riemsdijk, H. (Eds.), The Blackwell Companion to Syntax. Vol. 3. Blackwell.

Adamec, P., 1959. Pronominální podmět a významová výstavba věty v ruštině a v češtině. Bulletin VŠRJaL 3, 23–39.

Adamec, P., 1983. České zájmeno ten a jeho ruské ekvivalenty. In: Konfrontační studium ruské a české gramatiky a slovní zásoby 2. Universita Karlova, Praha.

Adamec, P., 1988. K vyjadřování a rozpoznávání koreference v ruštině a v češtině. In: Československá slavistika 1988. Praha, pp. 167–177.

Adams, M., 1987. From old french to the theory of pro-drop. Natural Language and Linguistic Theory 5, 1–32.

Alexiadou, A., Anagnostopoulou, E., 1998. Parameterizing Agr: Word Order, Verb-movement and EPP-checking. Natural Language and Linguistic Theory 16, 491–539.

Artstein, R., 1999. Person, animacy and null subjects. In: Cambier-Langeveld, T., Lipták, A., Redford, M., van der Torre, E. (Eds.), Proceedings of Console VII. SOLE, Leiden.

Auwera, J. v. d., 1984. Subject and non-Subject Asymmetries in the Relativization of Embedded NP's. In: de Gest, W., Putseys, Y. (Eds.), Sentential Complementation. Kluwer, Dordrecht, pp. 257–269.

Baayen, H., 2008. Analyzing Linguistic Data. Cambridge University Press.

Baayen, H. R., 2001. Word Frequency Distributions. Kluwer.

Bibliography

Baerman, M., Brown, D., 2011. Syncretism in Verbal Person/Number Marking, Ch. 29. In: Dryer and Haspelmath (2011).
URL http://wals.info

Bartnicka, B., Hansen, B., Klemm, W., Lehmann, V., Satkiewicz, H., 2004. Grammatik des Polnischen. Otto Sagner, München.

Bauer, J., 1967. K vývoji vztažných vět v slovanských jazycích. Slavica Slovaca 2 (2), 297–320.

Beesley, K., Karttunen, L., 2003. Finite-State Morphology. CSLI Studies in Computational Linguistics. The University of Chicago Press.

Berger, T., 1991. Zum denotativen Status "unbestimmt-persönlicher" Konstruktionen im Tschechischen. In: Grochowski, M., Weiss, D. (Eds.), Words are Physicians for an Ailing Mind. München, pp. 71–81.

Berger, T., 1993. Das System der tschechischen Demonstrativpronomina. Textgrammatische und stilspezifische Gebrauchsbedingungen. München, habilitationsschrift.

Berger, T., Weiss, D., 1987. Die Gebrauchsbedingungen des Anaphorikums "tot" in substantivischer Verwendung. In: Freidhof, G., Kosta, P. (Eds.), Slavistische Linguistik 1986. Sagner, München, pp. 9–93.

Biber, D., 1993. Representativeness in Corpus Design. Literary and Linguistic Computing 8 (4), 243–257.

Biberauer, T., 2010. Semi null-subject languages, expletives and expletive pro reconsidered. In: Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (Eds.), Parametric Variation: Null Subjects in Minimalist Theory. Cambridge University Press, Cambridge, pp. 154–199.

Bickel, B., 2003. Referential Density in Discourse and Syntactic Typology. Language 79, 708–736.

Blevins, J., 2003. Passives and Impersonals. Journal of Linguistics 39, 473–520.

Borer, H., 1986. I-subjects. Linguistic Inquiry 17, 375–416.

Borkovskij, V. I., 1949. Sintaksis drevnerusskich gramot. Prostoe predloženie. Izd.-vo L'vovskogo gosudarstvennogo universiteta, L'vov.

Borkovskij, V. I. (Ed.), 1968. Sravnitel'no-istoričeskij sintaksis vostočnoslavjanskich jazykov: tipy prostogo predloženija. Vol. 2. Nauka, Moskva.

Borkovskij, V. I. (Ed.), 1978. Istoričeskaja grammatika russkogo jazyka: Sintaksis. Prostoe predloženie. Nauka, Moskva.

Borkovskij, V. I., Kuznecov, P. S., 1965. Istoričeskaja grammatika russkogo jazyka, 2nd Edition. Nauka, Moskva.

Bräuer, H., 1952. Der persönliche Agens beim Passiv im Altbulgarischen. Eine syntaktische Untersuchung. Wiesbaden.

Brückner, A., 1998. Słownik etymologiczny języka polskiego. Wiedza Powszechna, Warszawa.

Burghardt, M., 2008. Annotationswerkzeuge für diachrone Korpora. Klassifikation und Evaluation von Annotationswerkzeugen. Master's thesis, Universität Regensburg.

Büring, D., 1999. Topic. In: Bosch, P., van der Sandt, R. (Eds.), Focus. Linguistic, cognitive and computational perspectives. Cambridge University Press, Cambridge, pp. 142–165.

Cabredo Hofherr, P., 2003. Arbitrary Readings of 3pl Pronominals. In: Weisgerber, M. (Ed.), Proceedings of the Conference "sub7 – Sinn und Bedeutung". Vol. 114 of Arbeitspapiere des Fachbereichs Sprachwissenschaft. Fachbereich Sprachwissenschaft der Universität Konstanz, Konstanz, pp. 81–94.

Cabredo Hofherr, P., 2006. 'Arbitrary" Pro and the theory of Pro-drop. In: Ackema, P., Brandt, P., Schoorlemmer, M., Weerman, F. (Eds.), Arguments and Agreement. Oxford University Press, Oxford, pp. 230–260.

Cardinaletti, A., Starke, M., 1999. The Typology of Structural Deficiency: A Case Study of Three Classes of Pronouns. In: van Riemsdijk, H. (Ed.), Clitic and Other Functional Categories in European Languages. Mouton de Gruyter, Berlin, pp. 145–233.

Chierchia, G., 1995. The variability of impersonal subjects. In: Bach, E., Jelinek, E., Kratzer, A., Partee, B. (Eds.), Quantification in Natural Languages. Vol. 1. Kluwer, pp. 107–143.

Chomsky, N., 1957. Syntactic Structures. Mouton, The Hague.

Chomsky, N., 1981. Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, N., 1982. Some Concepts and Consequences of the Theory of Government and Binding. MIT Press, Cambridge, Mass.

Bibliography

Chrakovskij, V. S., 1973. Passive Constructions. In: Kiefer, F. (Ed.), Trends in Soviet Linguistics. Reidel, Dordrecht, pp. 59–76.

Chrakovskij, V. S., 1974. Passivnye konstrukcii. In: Cholodovič, A. A. (Ed.), Tipologija passivnych konstrukcij. Izd. "Nauka", Leningrad, pp. 5–45.

Cinque, G., 1988. On *si* constructions and the theory of *arb*. Linguistic Inquiry 19, 521–581.

Claridge, C., 2008. Historical corpora. In: Lüdeling, A., Kytö, M. (Eds.), Korpuslinguistik/Corpus Linguistics. Vol. 2. DeGruyter, Berlin/New York, pp. 242–259.

Comrie, B., 1977. In defense of spontaneous demotion: The impersonal passive. In: Cole, P., Saddock, J. M. (Eds.), Grammatical Relations. New York, pp. 47–58.

Crystal, D., 1985. A dictionary of linguistics and phonetics, 2nd Edition. Basil Blackwell, London.

Cysouw, M., 2001. The paradigmatic structure of person marking. Ph.D. thesis, Katholieke Universiteit, Nijmegen, revised version published 2003: Oxford University Press.

D'Alessandro, R., 2006. Is Impersonal si in Italian Definite or Indefinite? In: Martínez, M. T., Alcazár, A., Hernández, R. M. (Eds.), Proceedings of the thirtythird Western Conference On Linguistics (WECOL 2004). UC Publications, Fresno, pp. 64–75.

Daneš, F., Grepl, M., Hlavsa, Z. (Eds.), 1987. Mluvnice češtiny. Vol. 3. Praha: Academia.

Diels, P., 1963. Altkirchenslavische Grammatik : mit einer Auswahl von Texten und einem Wörterbuch. Winter, Heidelberg.

Diessel, H., 1999. Demonstratives: Form, Function, and Grammaticalization. Vol. 42 of Typological Studies in Language. John Benjamins, Amsterdam.

Dipper, S., 2005. Xml-based stand-off representation and exploitation of multi-level linguistic annotation. In: Proceedings of Berliner XML Tage 2005 (BXML 2005). Berlin, Germany, pp. 39–50.

Dipper, S., Götze, M., Stede, M., 2004. Simple Annotation Tools for Complex Annotation Tasks: an Evaluation. In: Proceedings of the LREC Workshop on XML-based Richly Annotated Corpora. Lisbon, Portugal, pp. 54–62.

Dryer, M. S., 2011. Expression of Pronominal Subjects, Ch. 101. In: Dryer and Haspelmath (2011).
URL http://wals.info

Dryer, M. S., Haspelmath, M. (Eds.), 2011. World Atlas of Language Structures Online. Max Planck Digital Library, Munich.
URL http://wals.info

Dyła, S., 1983. A Non-Argument Against Multiple COMP Polish. Wiener Linguistische Gazette 30, 19–26.

D'Alessandro, R., Alexiadou, A., 2002. Inclusive and exclusive impersonal pronouns: a feature-geometrical analysis. Rivista di Grammatica Generativa 27, 31–44.

Eckhoff, H. M., Meyer, R., 2011. Conditions on Null Subjects in Old Church Slavonic, a Contrastive View, presentation at the 20. International Conference on Historical Linguistics, Osaka, 25.-30.07.2011.
URL   http://www.uni-regensburg.de/sprache-literatur-kultur/slavistik/institut/ sprach-und-kulturwissenschaft/roland-meyer/eckhoff_meyer_2011_slides_final.pdf

Engel, U., Rytel-Kuc, D., 1999. Deutsch-polnische kontrastive Grammatik. Julis Groos Verlag, Hamburg.

Evert, S., 2006. How Random is a Corpus? The Library Metaphor. Zeitschrift für Anglistik und Amerikanistik 54, 177–190.

Fassi Fehri, A., 2009. Arabic Silent Pronouns, Person, and Voice. Brill's Annual of Afroasiatic Languages and Linguistics 1, 3–40.

Fehrmann, D., Junghanns, U., 2008. Subjects and Scales. In: Richards, M., Malchukov, A. L. (Eds.), Scales. Vol. 86 of Linguistische Arbeitsberichte. Institut für Linguistik der Universität Leipzig, Leipzig, pp. 189–220.

Fehrmann, D., Junghanns, U., Lenertová, D., 2010. Two reflexive markers in slavic. Russian Linguistics 34, 203–238.

Filimonova, E. (Ed.), 2005. Clusivity. Typology and case studies of the inclusive–exclusive distinction. Vol. 63 of Typological Studies in Language. John Benjamins, Amsterdam, Philadelphia.

Fox, B. A., 1987. Discourse structure and anaphora: Written and conversational English. Cambridge University Press, Cambridge.

Franks, S., 1995. Parameters of Slavic Morphosyntax. Oxford University Press, New York.

Franz, N., Tuschinsky, C., 1982. Lektürehilfe für altrussische Texte. 11.-17. Jahrhundert. Vol. 1 of Mainzer Slavistische Veröffentlichungen - Slavica Moguntiaca. Liber Verlag, Mainz.

Frascarelli, M., 2007. Subjects, topics and the interpretation of referential pro. An interface approach to the linking of null pronouns. Natural Language and Linguistic Theory 25, 691–734.

Frascarelli, M., Hinterhölzl, R., 2007. Types of topics in German and Italian. In: Winkler, S., Schwabe, K. (Eds.), On information structure, meaning and form. John Benjamins, Amsterdam / Philadelphia, p. 87–116.

Fuß, E., to appear. Some remarks on the historical origin of pro-drop. In: Gallmann, P., Wratil, M. (Eds.), Null Pronouns. de Gruyter, Berlin.

Gau, M., 2005. The State of Historical Corpus Linguistics, with Special Focus on the Russian Language. Master's thesis, Universität Regensburg, Regensburg, http://www-slavistik.uni-regensburg.de/diakorp/.

Gebauer, J., 1970. Slovník staročeský, 2nd Edition. Academia, Praha.

Gebauer, J., 2007. Historická mluvnice jazyka českého. Vol. 4. Academia, Praha.

Gelderen, E. v., 2000. A history of English reflexive pronouns: person, self, and interpretability. John Benjamins, Amsterdam.

Geniušienė, E., 1987. The typology of reflexives. Mouton de Gruyter, Berlin, New York, Amsterdam.

Georgieva, V. L., 1968. Istorija sintaksičeskich javlenij russkogo jazyka. Prosveščenie, Moskva.

Gilligan, G., 1987. A Cross-Linguistic Approach to the pro-Drop Parameter. Ph.D. thesis, USC, Los Angeles.

Givón, T., 1983. Topic continuity in discourse. A quantitative cross-language study. Typological Studies in Language. John Benjamins, Amsterdam.

Givón, T., 2001. Syntax. An introduction. Benjamins, Amsterdam.

Grenoble, L. A., 1998. Deixis and Information Packaging in Russian Discourse. John Benjamins, Amsterdam/Philadelphia.

Grepl, M., 1973. Deagentnost a pasívum v slovanských jazycích. In: Českoslvenské přednášky pro VII. mezinárodní sjezd slavistů . Praha, pp. 141–149.

Grewendorf, G., 1990. Small pro in German. In: Grewendorf, G., Sternefeld, W. (Eds.), Scrambling and Barriers. Benjamins, Amsterdam, pp. 295–315.

Gundel, Jeanette, N. H., Zacharski, R., 1993. Cognitive status and the form of referring expressions in discourse. Language 69, 274–307.

Górna, M., 1976. Word Order – A Semantic Phenomenon? Papers and Studies in Contrastive Linguistics 5, 201–217.

Haegeman, L., 1997. Register variation, truncation and subject omission in english and in french. English Language and Linguistics 1, 233–270.

Haegeman, L., Ihsane, T., 1999. Subject ellipsis in embedded clauses in english. English Language and Linguistics 3 (1), 117–145.

Haider, H., 2001. Parametrisierung in der Generativen Grammatik. In: Haspelmath, M., König, E., Oesterreicher, W., Raible, W. (Eds.), Language typology and language universals. Handbücher zur Sprach- und Kommunikationswissenschaft. De Gruyter, Berlin, pp. 283–293.

Haspelmath, Martin, M. S. D. D. G., Comrie, B. (Eds.), 2005. World atlas of linguistic structures. Oxford University Press, Oxford.

Haug, D. T. T., Jøhndal, M. L., 2008. Creating a parallel treebank of the Old Indo-European Bible translations. In: Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). Workshop on language technology for cultural heritage data. Paris, pp. 27–34.

Haug, D. T. T., et al., 2009. Breaking down and putting back together: analysis and synthesis of New Testament Greek. Journal of Greek Linguistics 9, 56–92.

Havránek, B., 1928. Genera verbi v slovanských jazycích. Vol. 1. Nakladatelství Kr. České Spol. Nauk, Praha.

Bibliography

Havránek, B., 1937. Genera verbi v slovanských jazycích. Vol. 2. Nakladatelství Kr. České Spol. Nauk, Praha.

Himmelmann, N. P., 1997. Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur. Niemeyer, Tübingen.

Hirschová, M., 1984. K jednomu typu výpovědí běžně mluvené češtiny. Slovo a slovesnost 45, 273–275.

Holmberg, A., 2005. Is There a Little Pro? Evidence from Finnish. Linguistic Inquiry 36 (4), 533–564.

Holmberg, A., 2010. Null Subject Parameters. In: Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (Eds.), Parametric Variation: Null Subjects in Minimalist Theory. Cambridge University Press, Cambridge, pp. 88–124.

Holmberg, A., Nikanne, U., 2002. Expletives, subjects, and topics in Finnish. In: Svenonius, P. (Ed.), Subjects, expletives, and the EPP. Oxford University Press, Oxford, pp. 71–106.

Holmberg, A., Sheehan, M., 2010. Control into finite clauses in partial null-subject languages. In: Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (Eds.), Parametric Variation: Null Subjects in Minimalist Theory. Cambridge University Press, Cambridge, pp. 125–152.

Householder, F. (Ed.), 1981. Syntax of Apollonius Dyscolus. John Benjamins, Amsterdam.

Huang, Y., 2000. Discourse anaphora: Four theoretical models. Journal of Pragmatics 32, 151–176.

Hudousková, A., 2010. Reflexive clitics in czech. Ph.D. thesis, Univerzita Palackého, Olomouc.

Issatschenko, A., 1980. Geschichte der russischen Sprache. 1. Band: Von den Anfängen bis zum Ende des 17. Jahrhunderts. Slavica. Sammlung slavischer Lehr- und Handbücher. Neue Folge. Winter, Heidelberg.

Istrina, S. E., 1923. Sintaksičeskie javlenija Sinodal'nogo spiska I-j Novgorodskoj letopisi. Petrograd.

Ivanov, V., 1990. Istoričeskaja grammatika russkogo jazyka, 3rd Edition. Prosveščenie, Moskva.

Jaeggli, O., Safir, K., 1989. The Null Subject Parameter and Parametric Theory. In: Jaeggli, O., Safir, K. (Eds.), The Null Subject Parameter. Kluwer, Dordrecht.

Jakobson, R., 1939 (1971). Signe zéro. In: Selected Writings. Vol. 2. The Hague, pp. 211–219.

Janke, G., 1960. Der Ausdruck des Passivs im Altrussischen. Akademie-Verlag, Berlin.

Karas, M., et al., 1977-2005. Słownik gwar polskich. Ossolineum, Wrocław.

Karlsson, F., 2008. Early generative linguistics and empirical methodology. In: Lüdeling, A., Kytö, M. (Eds.), Korpuslinguistik/Corpus Linguistics. Vol. 1 of Handbücher zur Sprach- und Kommunikationswissenschaft. De Gruyter, Berlin/New York, pp. 14–32.

Kaufmann, I., 2004. Medium und Reflexiv. Eine Studie zur Verbsemantik. Vol. 489 of Linguistische Arbeiten. Niemeyer, Tübingen.

Keenan, E. L., 1975. Towards a Universal Definition of 'Subject'. In: Li, C. N. (Ed.), Subject and Topic. Academic Press, New York, pp. 303–335.

Keipert, H., 1999. Geschichte der russischen Literatursprache. In: Jachnow, H. (Ed.), Handbuch der sprachwissenschaftlichen Russistik und ihrer Grenzdisziplinen. Harrassowitz, Wiesbaden, pp. 726–779.

Kibort, A., 2004. Passive and passive-like constructions in English and Polish. Ph.D. thesis, University of Cambridge.

Kibrik, A. A., 1996. Anaphora in Russian Narrative Prose: A Cognitive Calculative Account. In: Studies in Anaphora. John Benjamins, pp. 255–303.

Klemensiewicz, Z., 1999. Historia języka polskiego, 7th Edition. PWN, Warszawa.

Koktová, E., 1992. Anaphoric Expressions Across Clause Boundary in pro-drop Languages. Prague Bulletin of Mathematical Linguistics 58, 39–63.

Kosta, P., 1990. Leere Kategorien in den nordslavischen Sprachen. Zur Analyse leerer Subjekte und Objekte in der Rektions-Bindungs-Theorie. Frankfurt/Main, habilitationsschrift, Universität Frankfurt/Main.

Kratzer, A., 1997. German Impersonal Pronouns and Logophoricity, presentation, *Sinn und Bedeutung*, Berlin 1997 (handout at http://semanticsarchive.net/Archive/WViZTE1M/Impersonalogophoricity.pdf).

Kresin, S., 1994. Third Person Reference in Russian and Czech. Ph.D. thesis, University of California at Berkeley.

Bibliography

Kroch, A., 1989. Reflexes of grammar in patterns of language change. Language Variation and Change 1, 199–244.

Kroch, A., 2000. Syntactic Change. In: Baltin, M., Collins, C. (Eds.), Handbook of Contemporary Syntactic Theory. Blackwell, Oxford.

Krys'ko, V. B., 1995. Drevnerusskaja grammatika XII - XIII vv. Nauka, Moskva, Ch. Zalogovye otnošenija, pp. 465–506.

Krzek, M., 2010. Some Aspects of Subjects of Impersonal Constructions in Polish. Newcastle Working Papers in Linguistics 16, 66–87.

Kytö, M., 1996. Manual to the diachronic part of the Helsinki Corpus of English Texts: Coding conventions and lists of source texts. Tech. rep., Universität Helsinki, http://khnt.hit.uib.no/icame/manuals/HC/.

Kytö, M., Rissanen, M., 1993. General Introduction. Mouton de Gruyter, Berlin / New York, Ch. 1, pp. 1–17.

Křížková, H., 1974. K novějšímu pokusu o výklad vztahu zájmena ten a on v češtině. Naše řeč 57, 85–89.

Lehmann, C., 1995. Thoughts on grammaticalization. Lincom Europa, München, [= ms. 1982, Universität zu Köln].

Levinson, S. C., 1983. Pragmatics. Cambridge University Press.

Lichačev, D. S., et al. (Eds.), 2000. Biblioteka literatury Drevnej Rusi. Vol. 1. Russkaja Akademija Nauk, Institut Russkoj literatury, Moskva.

Lightfoot, D., 1979. Principles of Diachronic Syntax. Cambridge University Press, New York.

Linde, M. S. B., 1854-61. Słownik języka polskiego. Lwów.

Lindseth, M., 1994. Overt expletives and thematic subjects in West Slavic. Studies in the Linguistic Sciences 24 (2), 333–346.

Lindseth, M., 1998. Null-subject properties of Slavic languages: with special reference to Russian, Czech and Sorbian. Vol. 361 of Slavistische Beiträge. Sagner, München.

Lindseth, M., Franks, S., 1996. Licensing and Identification of Null Subjects in Slavic. In: Toman, J. (Ed.), Formal Approaches to Slavic Linguistics: The College Park Meeting. Slavica, pp. 199–224.

Lomonosov, M. V., 1755. Rossijskaja grammatika. Imperatorsjaka Akademija Nauk, Sankt Peterburg.

Lomtev, T. P., 1956. Očerki po istoričeskomu sintaksisu russkogo jazyka. Izd.-vo Moskovskogo universiteta, Moskva.

Malamud, S., 2007. Impersonal indexicals: *you* and *man, si.*, ms., Brandeis University, http://people.brandeis.edu/ smalamud/iiss.pdf.

Malamud, S. A., 2006. Semantics and Pragmatics of Arbitrariness. Ph.D. thesis, University of Pennsylvania.

Malamud, S. A., to appear a. (in)definiteness-driven typology of arbitrary items, http://people.brandeis.edu/ smalamud/malamud-vol.pdf.

Malamud, S. A., to appear b. Impersonal indexicals: *one, you, man*, and *du.* Journal of Comparative Germanic Linguisticshttp://people.brandeis.edu/ smalamud/malamud-JCGL-web-draft.pdf.

Marti, R., 1989. Handschrift – Text – Textgruppe – Literatur. Vol. 68 of Osteuropa-Institut an der Freien Universität Berlin – Slavistische Veröffentlichungen. Otto Harrassowitz, Wiesbaden.

Marušič, F., Žaucer, R., 2006. On the intensional *feel-like* construction in Slovenian: A case of a phonologically null verb. Natural Language and Linguistic Theory 24, 1093–1159.

Mayol, L., to appear. An account of the variation in the rates of overt subject pronouns in Romance, in: *Spanish in Context.* Available from http://laiamayol.wordpress.com/publications/.

McEnery, T., Wilson, A., 2002. Corpus Linguistics, 2nd Edition. Edinburgh University Press, Edinburgh.

McGann, W., Schwarz, A., 1988. Main character in children's narrative. Linguistics 26, 215–233.

McShane, M., 2009. Subject Ellipsis in Russian and Polish. Studia Linguistica 63 (1), 98–132.

Bibliography

Mel'čuk, I., 1974. O sintaksičeskom nule. In: Cholodovič, A. A. (Ed.), Tipologija passivnych konstrukcij, diatezy i zaloga. Leningrad, p. 343–361.

Meyer, R., 2004. Syntax der Ergänzungsfrage. Empirische Untersuchungen am Russischen, Polnischen und Tschechischen. Vol. 436 of Slavistische Beiträge. München.

Meyer, R., 2005. The Regensburg Diachronic Corpus of Russian: A new source for linguistic research (not only) on modality. In: Hansen, B., Karlík, P. (Eds.), Modality in Slavonic Languages: New Perspectives. Verlag Otto Sagner, München, pp. 315–338.

Meyer, R., 2009a. Interjection > pronoun? The history of Czech expletive on(o). In: Karlík, P. (Ed.), Development of Language through the Lens of Formal Linguistics. No. 43 in Lincom Studies in Theoretical Linguistics. Lincom Europe.

Meyer, R., 2009b. Zur Geschichte des referentiellen Nullsubjekts im Russischen. Zeitschrift für Slavistik 54, 375–397, ms., University of Regensburg.

Meyer, R., 2010. Reflexive passives and impersonals in north slavonic languages: a diachronic view. Russian Linguistics 34 (3), 285–306.

Meyer, R., 2011. Old wine in new wineskins? Tagging Old Russian via annotation projection from modern translations. Russian Linguistics 35 (2), 267–281.

Moltmann, F., 2006. Generic one, arbitrary PRO, and the first person. Natural Language Semantics 14, 257–281.

Müller, G., 1988. Zur Analyse subjektloser Konstruktionen in der Rektions-Bindungs-Theorie. That-trace-Effekte und pro-Lizensierung im Russischen, universität Frankfurt/Main.

Müller, G., 1993. On Deriving Movement Type Asymmetries. Ph.D. thesis, Universität Tübingen.

Müller, G., 2006. Pro-Drop and Impoverishment. In: Brandt, P., Fuß, E. (Eds.), Form, Structure, and Grammar. A Festschrift Presented to Günther Grewendorf on Occasion of his 60th Birthday. Akademie-Verlag, Berlin, pp. 93–115, university of Leipzig.

Müller, G., 2007. Some consequences of an impoverishment-based approach to morphological richness and pro-drop, ms., University of Leipzig.
URL http://www.uni-leipzig.de/~muellerg/mu228.pdf

Müller, G., Rohrbacher, B., 1989. Eine Geschichte ohne Subjekt. Zur Entwicklung der pro-Theorie. Linguistische Berichte 119, 3–52.

Neeleman, A., Szendrői, K., 2007. Radical Pro Drop and the Morphology of Pronouns. Linguistic Inquiry 38 (4), 671–714.

Newmeyer, F., 2001. Deconstructing grammaticalization. Language Sciences 23, 187–229.

Newmeyer, F. J., 2004. Against a parameter-setting approach to typological variation. Linguistic Variation Yearbook 4, 181–234.

Newmeyer, F. J., 2006. A rejoinder to 'On the role of parameters in Universal Grammar: a reply to Newmeyer' by Ian Roberts and Anders Holmberg, ms., University of Washington. Revised version, February 2006.

Nichols, J., 1984. The Grammatical Marking of Theme in Literary Russian. In: Brecht, R. D., Flier, M. S. (Eds.), Issues in Russian Morphosyntax. Slavica, Columbus.

Nilsson, B., 1982. Personal pronouns in Russian and Polish : a study of their communicative function and placement in the sentence. Stockholms Universitet: [Acta Universitatis Stockholmiensis / Stockholm Slavic studies] ; 13. Almqvist & Wiksell, Stockholm.

Oliva, K., 2001. Reflexe reflexivity reflexiv. Slovo a slovesnost 57, 200–207.

Padučeva, E. A., 1985. Vyskazyvanie i ego sootnosennost' s dejstvitel'nost'ju. Nauka, Moskva.

Padučeva, E. V., 1981. Mestoimenie 'èto' s predmetnym antecedentom. In: Problemy strukturnoj lingvistiki, 1979. Akademija nauk, Moskva, pp. 72–88.

Pepłowski, F., 1966-2004. Słownik polszczyzny XVI wieku. Wydawnictwo Polskiej Akademii Nauk, Wrocław.

Pereira, F., 2000. Formal grammar and information theory: together again? Philosophical Transactions of the Royal Society 358, 1239–1253.

Perlmutter, D., 1971. Deep and surface structure constraints in syntax. Holt, Rinehart, and Winston, New York.

Pesetsky, D., 1987. Wh-in-Situ: Movement and Unselective Binding. In: Reuland, E., ter Meulen, A. (Eds.), The Representation of (In)Definiteness. MIT Press, Cambridge, Mass., pp. 98–129.

Bibliography

Pisarkowa, K., 1984. Historia składni języka polskiego. Ossolineum, Wrocław.

Remneva, M. L., 2003. Puti razvitija russkogo literaturnogo jazyka XI-XVII vv. Učebnoe posobie po kursu "Istorija russkogo literaturnogo jazyka". Izd.-vo Moskovskogo universiteta, Moskva.

Rezac, M., 2004. Elements of cyclic syntax: Agree and Merge. Ph.D. thesis, University of Toronto.

Ribarov, K., Bubník, J., Čelák, J., Vojtěch, J., Kára, A., Novák, V., Vondra, T., 2004. ACT - Computer Processing of Written Cultural Heritage Sources. In: INFORUM 2004: 10th Conference on Professional Information Resources. Prague, May 25-27, 2004.

Rijksbaron, A., 1984. The Syntax and Semantics of the Verb in Classical Greek. An Introduction. Grieben, Amsterdam.

Rissanen, M., 1989. Three problems connected with the use of diachronic corpora. ICAME Journal 13, 16–19.

Rissanen, M., 2008. Corpus linguistics and historical linguistics. Vol. 1 of Handbücher zur Sprach- und Kommunikationswissenschaft. De Gruyter, Berlin/New York, pp. 53–68.

Rivero, M. L., Milojević-Sheppard, M., 2003. Indefinite Reflexive Clitics in Slavic: Polish and Slovenian. Natural Language and Linguistic Theory 21, 89–155.

Rizzi, L., 1982. Issues in Italian syntax. Foris, Dordrecht.

Rizzi, L., 1986. Null Objects in Italian and the Theory of *pro*. Linguistic Inquiry 17, 501–557.

Rizzi, L., 1997. The Fine Structure of the Left Periphery. In: Haegeman, L. (Ed.), Elements of Grammar. Kluwer, Dordrecht, pp. 281–337.

Roberts, I., 1985. Agreement Parameters and the Development of the English Modal Auxiliaries. Natural Language and Linguistic Theory 3, 21–58.

Roberts, I., 2007a. Diachronic Syntax. Vol. 8 of Oxford Textbooks in Linguistics. Oxford University Press, Oxford, UK.

Roberts, I., Holmberg, A., 2010. Introduction: Parameters in minimalist theory. In: Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (Eds.), Parametric Variation: Null Subjects in Minimalist Theory. Cambridge University Press, pp. 1–57.

Roberts, I. G., 2007b. Taraldsen's Generalisation and Language Change: Two Ways to Lose Null Subjects, ms., Cambridge.
URL http://people.pwf.cam.ac.uk/mtb23/NSP/

Roberts, I. G., Holmberg, A., 2006. On the role of parameters in Universal Grammar: a reply to Newmeyer. In: Broekhuis, H., Corver, N., Huybregts, R., Kleinhenz, U., Koster, J. (Eds.), Organizing Grammar. Linguistic Studies in Honor of Henk van Riemsdijk. Mouton de Gruyter, Berlin, pp. 538–553.

Ross, J. R., 1982. Pronoun deleting processes in German, paper presented at the Annual Meeting of the Linguistics Society of America, San Diego, California.

Rothe, H., 2000. Was ist "altrussische Literatur"? Westdeutscher Verlag, Wiesbaden.

Rothstein, R. A., 1993. Polish. In: Comrie, B., Corbett, G. G. (Eds.), The Slavonic languages. Routledge, London, pp. 666–758.

Rytel-Kuc, D., 1990. Niemieckie passivum i man-Sätze a ich przekład w języku polskim i czeskim. Zakład narodowy im. Ossolińskich.

Růžička, R., 1986a. Funkcionirovanie i klassifikacija pustych kategorij v russkom literaturnom jazyke. Zeitschrift für Slavistik 31, 388–392.

Růžička, R., 1986b. Typologie der Diathese slavischer Sprachen in parametrischen Variationen. Die Welt der Slaven 21, 225–274.

Sadnik-Aitzetmüller, L., Aitzetmüller, R., 1955. Handwörterbuch zu den altkirchenslavischen Texten. Indogermanische Bibliothek : 2, Wörterbücher. Winter [u.a.], Heidelberg.

Schmalstieg, W. R., 1995. An Introduction to Old Russian. Vol. 15 of Journal of Indo-European Studies Monograph Series. Washington D.C.

Schmidt, W.-H., 2002. Mittelalter. In: Städtke, K. (Ed.), Russische Literaturgeschichte. Metzler, Stuttgart, pp. 1–62.

Schäfer, F., 1982. Untersuchungen zur Reflexivität im Alttschechischen. Beginn des Schrifttums bis 1350. Vol. 52 of Veröffentlichungen der Abteilung für Slavische Sprachen und Literaturen des Osteuropa-Instituts (Slavisches Seminar) an der Freien Universität Berlin. Harrassowitz.

Bibliography

Selig, M., 1998. Pseudoreflexivität im Altitalienischen – Voraussetzungen und Richtungen eines Grammatikalisierungsprozesses. In: Geisler, H., Jacob, D. (Eds.), Transitivität und Diathese in romanichen Sprachen. Niemeyer, Tübingen.

Siekierska, K. (Ed.), 1999-2004. Słownik języka polskiego XVII i 1. połowy XVIII wieku. Wydawn. Inst. Języka Polskiego PAN, Kraków.

Siewierska, A., 1988. The Passive in Slavic. In: Shibatani, M. (Ed.), Passive and Voice. John Benjamins, Amsterdam, p. 243–289.

Siewierska, A., 2008a. Introduction: Impersonalization from a subject-centred vs. agent-centred perspective. Transactions of the Philological Society 106 (2), 1–23.

Siewierska, A., 2008b. Ways of impersonalizing: Pronominal vs. verbal strategies. In: de los Angeles, G., Gonzales, M., Mackenzie, L., Gonzalez Alvarez, E. (Eds.), Current Trends in Contrastive Linguistics. John Benjamins, Amsterdam, pp. 3-26. John Benjamins, pp. 3–26.

Sigurðsson, H., Egerland, V., 2009. Impersonal Null Subjects in Icelandic and Elsewhere. Studia Linguistica 63 (1), 158–185.

Silberztein, M., 2000. INTEX : an FST toolbox. Theoretical computer science 231 (1), 33–46.

Silberztein, M., 2005. NooJ's Dictionaries. In: Proceedings of LTC 2005. Poznań.

Sreznevskij, I. I., 1955. Materialy dlja slovarja drevne-russkago jazyka po pis'mennym pamjatnikam. Vol. 1-3. Akad. Druck- und Verlagsanstalt, Graz.

Stefanowitsch, A., 2006. Negative evidence and the raw frequency fallacy. Corpus Linguistics and Linguistic Theory 2, 61–77.

Stoll, S., Bickel, B., 2009. How deep are differences in referential density? In: Crosslinguistic approaches to the Psychology of Language. Research in the Tradition of Dan Isaac Slobin. Psychology Press, pp. 543–556.

Szucsich, L., 2008. Evidenz für syntaktische Nullen aus dem Burgenlandkroatischen, Polnischen, Russischen und Slovenischen: Merkmalsausstattung, Merkmalshierarchien und morphologische Defaults. Zeitschrift für Slavistik 53 (2), 160–177.

Taraldsen, K. T., 1978. On the NIC, vacuous application and the that-trace filter, mIT, Indiana Linguistics Club.

Tiedemann, J., 2004. Word to word alignment strategies. In: Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004). Geneva, Switzerland, pp. 212–218.

Trávníček, F., 1930. Neslovesné věty v češtině. Vol. 1 – Věty interjekční. Filozofická fakulta MU, Brno.

Trávníček, F., 1939. Pasívum ve spisovné češtině. Slovo a slovesnost 5, 13–24.

Trávníček, F., 1951. Mluvnice spisovné češtiny. Vol. 2. Slovanské nakladatelství, Praha.

Trávníček, F., 1961. Historická mluvnice česká 3: Skladba. Státní pedagogické nakladatelství, Praha.

Uhlířová, L., 1987. O (ne)vyjadřování zájmenného podmětu 3. osoby v horní lužické srbštině v porovnání s češtinou. Lětopis institutu za serbski ludospyt, Rjad A – rěčespyt 34, 19–30.

Unbegaun, B. O. (Ed.), 1959. Henrici Wilhelmi Ludolfi Grammatica Russica. Oxonii A. D. MDCXCVI. Clarendon Press.

Urbańczyk, S. (Ed.), 1953-2002. Słownik staropolski. PAN, Warszawa.

Vainikka, A., Levy, Y., 1999. Empty subjects in Hebrew and Finnish. Natural Language and Linguistic Theory 17, 613–671.

Van Valin, R. D. J., LaPolla, R. J., 1997. Syntax. Structure, meaning and function. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.

Večerka, R., 1996. Altkirchenslavische (Altbulgarische) Syntax. III. Die Satztypen: Der einfache Satz. Vol. 27 of Monumenta linguae Slavicae dialecti veteris. Weiher, Freiburg i. Br.

Weiss, D., 1977. Syntax und Semantik polnischer Partizipialkonstruktionen im Rahmen einer generativ-transformationellen Sprachbeschreibung. Vol. 10 of Slavica Helvetica. Bern, Frankfurt/Main.

Weiss, D., 1993. Die Faszination der Leere. Die moderne russische Umgangssprache und ihre Liebe zur Null. In: Zeitschrift für Slavische Philologie LIII/1993 (Beiträge zum XI. Internationalen Slavisten-Kongreß, Bratislava 1993). pp. 48–82.

Wiemer, B., to appear. On the rise, establishment and continued development of subject impersonals in Polish, East Slavic and Baltic, ms., Universität Konstanz.

Bibliography

Wiemer, B., Giger, M., 2005. Resultativa in den nordslavischen und baltischen Sprachen (Bestandsaufnahme unter arealen und grammatikalisierungstheoretischen Gesichtspunkten). Vol. 10 of LINCOM Studies in Language Typology. LINCOM Europa, München, Newcastle.

Wörner, K., Witt, A., Rehm, G., Dipper, S., 2006. Modelling linguistic data structures. In: Proceedings of the Extreme Markup Languages 2006. Montréal, Canada.

Zaliznjak, A. A., 2004. Drevnenovgorodskij dialekt., 2nd Edition. Vol. 1. Jazyki Slavjanskoj kul'tury, Moskva.

Zaliznjak, A. A., 2008. Drevnerusskie ènklitiki. Jazyki Slavjanskich Kul'tur, Moskva.

Zeldes, A., Ritz, J., Lüdeling, A., Chiarcos, C., 2009. Annis: A search tool for multi-layer annotated corpora. In: Proceedings of Corpus Linguistics 2009 July 20-23, Liverpool.

Zolotova, G. A., 1991. Sub"ektnye modifikacii russkogo predloženija. In: Words are physicians for an ailing mind. For Andrzej Bogusławski on the occasion of his 60th birthday. Maciej Grochowski and Daniel Weiss, München, pp. 509–515.

Zubatý, J., 1909. O jistém způsobu užívání zájmen an a on a jeho původu. Listy filologické 36, 25–34, 106–117, 230–240, 334–352.

Černych, P. J., 1954. Istoričeskaja grammatika russkogo jazyka. Učpedgiz, Moskva.

Štícha, F., 1985. Poznámky o staročeském pasívu. Listy fiologické 108, 77–82.

Štícha, F., 1988. K vývoji zvratného pasíva ve spisovné češtině. Listy filologické 111.