AGILE

Automatic Generation of Instructions in Languages of Eastern Europe

---

Title ***Formal specification of extended grammar models***

Authors Elena Andonova
John Bateman
Nevena Gromova
Anthony Hartley
Geert-Jan M. Kruijff
Ivana Kruijff-Korbayová
Serge Sharoff
Hana Skoumalová
Lena Sokolova
Kamenka Staykova
Elke Teich

Deliverables *WP6: SPEC2-Bu, SPEC2-Cz, SPEC2-Ru*

Status *Final, due month 14 (February 1999)*

Availability *Public*

Date *1999-02-28*

INCO COPERNICUS PL961104

Abstract:

This document describes the deliverables SPEC2-Bu, SPEC2-Cz and SPEC2-Ru of work package 6, task 6.2 of the AGILE project. According to the Technical Annexe, page 31, the objective for this workpackage consists in the specification of an extended grammar model enabling generation of the text for complex procedures in a single style. The task WP6.2 provides an extension of the phenomena for simple procedures researched by the task WP6.1.

In the introduction to this delivertable, we discuss the scope of the grammar to be modelled for target languages, present the methodology for the description of phenomena and a formal notation for their specification using notions of systemic-functional grammar (SFG). In Section 2, we present the following list of functional regions identified in our study and described in this deliverable: Transitivity, Diathesis, Mood, Tense, Aspect, Clause complexity, Word order and Determination in nominal groups. The description of each functional region starts with an introduction into SFG-based principles for its description, which is followed by treatments of respective phenomena in Bulgarian, Czech and Russian. The description of each phenonemon is finished with a cross-linguistic summary description comparing phenomena of target languages. Per-language treatments of functional regions include the presentation of basic phenomena available in a target language, the detailed description of phenomena which are present in our corpus, and their formal specifications in the proposed notation. The final section concludes the report by drawing comparisons from described functional regions in the target Slavonic languages, summarizing accomplishments of the presented research and suggestions for further research, and describing the implications of this study for implementation of the intermediate demonstrator.

More information on AGILE is available on the project web page and from the project coordinators:

URL:	http://www.itri.brighton.ac.uk/projects/agile
email:	agile-coord@itri.bton.ac.uk
telephone:	+44-1273-642900
f	+44-1273-642908

Table of Content

1. Introduction 1
1.1 Basic notions of systemic-functional linguistics 3
1.2 The AGILE approach to describing functional regions of Slavonic languages 4
1.3 Corpus investigation principles 5
1.4 A notation for specification of grammatical resources 6
2. The chosen phenomena in the Slavonic languages 8
2.1 Transitivity 8
2.1.1 Theoretical background 8
2.1.2 Description for Russian 9
2.1.2.1 Types of Processes 9
2.1.2.2 Types of Participants 13
2.1.2.3 Formal specifications of the nuclear transitivity 14
2.1.2.4 Types of Circumstances 16
2.1.2.4.1 Description of prepositional phrases 16
2.1.2.4.2 Formal specifications for prepositional phrases 20
2.1.2.4.3 Description of adverbial groups 24
2.1.3 Description for Bulgarian 24
2.1.3.1 Types of Processes 24
2.1.3.2 Types of Participants 27
2.1.3.3 Formal specifications of the nuclear transitivity 28
2.1.3.4 Types of Circumstances 29
2.1.3.4.1 Description of prepositional phrases 29
2.1.3.4.2 Formal specifications for prepositional phrases 30
2.1.3.4.3 Description of adverbial groups 31
2.1.4 Description for Czech 32
2.1.4.1 Types of Processes 32
2.1.4.1.1 Evidence from the corpus 32
2.1.4.1.2 Observations in terms of valency frames 39
2.1.4.2 Availability of Relevant Concepts in the UM 40
2.1.4.3 Proposals 41
2.1.4.4 Types of Participants and Circumstances 42
2.1.4.4.1 Formal specifications of the nuclear transitivity 43

2.1.4.4.2 Types of Prepositional Phrases in the corpus 44
2.1.5 Crosslinguistic summary description 46
2.2 Diathesis 47
2.2.1 Description for Russian 49
2.2.1.1 Theoretical background 49
2.2.1.2 Evidence from the corpus 50
2.2.1.3 Formal specifications 51
2.2.2 Description for Bulgarian 52
2.2.2.1 Theoretical background 52
2.2.2.2 Evidence from the corpus 52
2.2.3 Description for Czech 53
2.2.3.1 Theoretical background 54
2.2.3.1.1 Impersonal verbs 56
2.2.3.1.2 Personal verbs 56
2.2.3.1.3 Realisations of Actor in Czech 60
2.2.3.2 Evidence from the corpus 62
2.2.3.2.1 Active constructions 62
2.2.3.2.2 Passive constructions. 63
2.2.3.3 Formal specifications 64
2.2.4 Crosslinguistic summary description 65
2.3 Mood 68
2.3.1 Description for Bulgarian 70
2.3.1.1 Theoretical background 70
2.3.1.2 Evidence from the corpus 70
2.3.2 Description for Russian 71
2.3.2.1 Theoretical background 71
2.3.2.2 Evidence from the corpus 71
2.3.3 Description for Czech 72
2.3.3.1 Evidence from the corpus 72
2.3.4 Formal specifications 73
2.3.4.1 Indicative Clauses 73
2.3.4.2 Imperative Clauses 73
2.3.4.2.1 Formal specifications for Bulgarian 73
2.3.4.2.2 Formal specifications for Czech 74
2.3.4.2.3 Formal specifications for Russian 75

2.4 Tense 75
2.4.1 Theoretical background 75
2.4.2 Evidence from the corpus 77
2.4.3 Formal specifications 79
2.4.3.1 Formal specifications for Russian 79
2.4.3.2 Formal specifications for Bulgarian 79
2.4.3.3 Formal specifications for Czech 79
2.5 Aspect 79
2.5.1 Theoretical background 79
2.5.2 Evidence from the corpus 82
2.5.3 Formal specifications 83
2.5.3.1 Bulgarian 83
2.5.3.2 Czech 84
2.5.3.3 Russian 84
2.6 Clause Complexity 84
2.6.1 Introduction 84
2.6.2 Clause Complexity in SFG 85
2.6.2.1 Interdependency or Taxis between Clause Complexes 85
2.6.2.1.1 Hypotaxis 86
2.6.2.1.2 Parataxis 86
2.6.2.2 Logico-Semantic Relations between Clause Complexes 87
2.6.2.2.1 Expansion - Elaboration 87
2.6.2.2.2 Expansion - Extension 88
2.6.2.2.3 Expansion - Enhancement 89
2.6.3 Clause Complexity in Nigel 89
2.6.4 Evidence from the Corpus 93
2.6.4.1 Introduction 93
2.6.4.2 Corpus Data 93
2.6.4.2.1 Elaboration + Parataxis 93
2.6.4.2.2 Exposition 93
2.6.4.2.3 Elaboration + Hypotaxis 94
2.6.4.2.4 Extension + Parataxis 97
2.6.4.2.5 Extension + Hypotaxis 98
2.6.4.2.6 Enhancement + Parataxis 98
2.6.4.2.7 Enhancement + Hypotaxis 99

2.6.4.2.8 Some Connectives ... 103
2.6.5 Conclusions ... 103
2.7 Word Order ... 104
2.7.1 Description of phenomena ... 106
2.7.2 Thematic Structure and Information Structure in SFG ... 116
2.7.2.1 Thematic Structure in SFG ... 116
2.7.2.2 Information Structure in SFG ... 119
2.7.2.3 Word Order in SFG ... 119
2.7.3 Information Structure in FGD ... 121
2.7.3.1 Systemic Ordering and Communicative Dynamism ... 122
2.7.3.2 Contextual Boundness ... 124
2.7.3.3 Topic-Focus Articulation ... 125
2.7.3.4 TFA and Word Order ... 126
2.7.4 An Abstract Algorithm for Word Order ... 127
2.7.4.1 Assumptions and Basic Ideas ... 127
2.7.4.2 The Ordering Algorithm ... 129
2.7.5 Conclusions ... 129
2.8 Determination ... 130
2.8.1 Introduction of Terminology ... 132
2.8.2 Determination in the Corpora ... 137
2.8.3 Specific Determination ... 139
2.8.3.1 Specific Deictic Items in Czech ... 139
2.8.3.1.1 Evidence from the Czech Corpus ... 141
2.8.3.1.2 Formalization of Specific Deictics in Czech ... 142
2.8.3.2 Specific Deictic Items in Russian ... 143
2.8.3.2.1 Evidence from the Russian Corpus ... 143
2.8.3.2.2 Formalization of Specific Deictics in Russian ... 144
2.8.3.3 Specific Deictic Items in Bulgarian ... 145
2.8.3.3.1 Evidence from the Bulgarian Corpus ... 145
2.8.3.3.2 Formalization of Specific Deictics in Bulgarian ... 146
2.8.4 Partial Non-Specific Determination ... 146
2.8.4.1 Partial Existential determination in Czech ... 147
2.8.4.1.1 Evidence from the Czech Corpus ... 150
2.8.4.1.2 Formalization of Partial Existential Deictics in Czech ... 151
2.8.4.2 Partial Existential determination in Russian ... 151

2.8.4.2.1 Evidence from the Russian Corpus 152
2.8.4.2.2 Formalization of Partial Existential Deictics in Russian 152
2.8.4.3 Partial Non-Specific determination in Bulgarian 153
2.8.4.3.1 Evidence from the Bulgarian Corpus 153
2.8.4.3.2 Formalization of Partial Non-Specific Deictics in Bulgarian 154
2.8.4.4 Partial Variable determination in Czech 154
2.8.4.4.1 Evidence from the Czech Corpus 156
2.8.4.4.2 Formalization of Partial Variable Deictics in Czech 156
2.8.4.5 Partial Variable determination in Russian 156
2.8.4.5.1 Evidence from the Russian Corpus 156
2.8.4.5.2 Formalization of Partial Variable Deictics in Russian 157
2.8.5 Total Non-Specific Determination 157
2.8.5.1 Total Non-Specific Detemination in Czech 157
2.8.5.1.1 Evidence from the Czech Corpus 157
2.8.5.1.2 Formalization of Total Non-Specific Deictics in Czech 158
2.8.5.2 Total Non-Specific Detemination in Russian 158
2.8.5.2.1 Evidence from the Russian Corpus 158
2.8.5.2.2 Formalization of Total Non-Specific Deictics in Russian 159
2.8.5.3 Total Non-Specific Detemination in Bulgarian 159
2.8.5.3.1 Evidence from the Bulgarian Corpus 160
2.8.5.3.2 Formalization of Total Non-Specific Deictics in Bulgarian 160
2.8.6 Concluding Remarks 160
3. Conclusions and Further Work 162
4. Appendix 169
Bulgarian 169
Czech 169
Russian 170

Table of Figures

Figure 1 Russian: Relational processes 10
Figure 2 Russian: Dispositive-material-processes 12
Figure 3 Russian: Non-directed material processes 13
Figure 4 Russian: Adverbial groups 24
Figure 5 Bulgarian: Process types 27
Figure 6 Czech: creative directed material processes 36
Figure 7 Czech: dispositive directed material processes 37
Figure 8 Czech: Directed material processes requiring a Spatial-locative complement 38
Figure 9 Czech: Directed material processes requiring a Spatial-locative complement 38
Figure 10 Material actions in the UM 39
Figure 11 Cause-Effect in the UM 40
Figure 12 Initial model of TRANSITIVITY for Czech 43
Figure 13 Initial model of phase type for Czech 44
Figure 14 Initial formalization of the Czech voice system 65
Figure 15 Clause nexus / tactic dimensions 86
Figure 16 Types of extension 88
Figure 17 Types of enhancement 89
Figure 18 Connectives 103
Figure 19 Template sentence structure regarding informativity 128
Figure 20 Abstract algorithm for word order 129
Figure 21 Insertion of a deictic element 131
Figure 22 Items functioning as specific Deictic in English 132
Figure 23 Items functioning as non-specific Deictic in English 133
Figure 24 Realisation of a deictic element by an empty string 135
Figure 25 Explicit versus implicit deictic element 135
Figure 26 Corpus evidence of occurrences of determination and quantification 138
Figure 27 Items functioning as specific Deictic in Czech 141
Figure 28 Occurrences of specific Deictic items in the Czech corpus 141
Figure 29 Formal specification of Czech specific determination 143
Figure 30 Items functioning as specific Deictic in Russian 143
Figure 31 Occurrences of specific (and possessive) Deictic items in the Russian corpus 144
Figure 32 Formal specification of Russian specific determination 144

Figure 33 Items functioning as specific Deictic in Bulgarian 145
Figure 34 Occurrences of specific Deictic items in the Bulgarian corpus 145
Figure 35 Formal specification of Bulgarian specific determination 146
Figure 36 Formalization of the distinction between existential and variable non-specific determination 147
Figure 37 Formalization of the distinction between partial and total non-specific determination 147
Figure 38 Classification of Czech deictics according to knowledge of individual characterictic 148
Figure 39 Partial Non-Specific Existential Deictic items in Czech 150
Figure 40 Occurrences of nominal groups with an explicit existential determiner in the Czech corpus 150
Figure 41 Formalization of existential partial non-specific determination 151
Figure 42 Partial Non-Specific Existential Deictic items in Russian 152
Figure 43 Occurrences of nominal groups with an explicit existential determiner in the Russian corpus 152
Figure 44 Formalization of existential partial non-specific determination 152
Figure 45. Partial Non-Specific Deictics in Bulgarian 153
Figure 46. Formalization of the partial non-specific determination in Bulgarian 154
Figure 47 Partial non-specific variable deictic items in Czech 156
Figure 48 Formalization of existential non-specific determination 156
Figure 49 Formalization of existential non-specific determination 157
Figure 50 Total non-specific deictic items in Czech 157
Figure 51 Occurrences of nominal groups with an explicit total non-specific deictic 158
Figure 52 Formalization of total non-specific determination 158
Figure 53 Occurrences of nominal groups with an explicit total non-specific deictic 159
Figure 54 Formalization of total non-specific determination 159
Figure 55. Total Non-Specific Deictics in Bulgarian 159
Figure 56. Formalization of total non-specific determination in Bulgarian 160

1. Introduction

The primary goal of AGILE consists in developing a suite of software tools to assist technical writers in the production of CAD/CAM manuals in selected languages of Eastern Europe: Bulgarian, Czech and Russian. This problem is approached by means of multilingual generation from a common semantic representation of the procedural aspects of the task of using such software tools. Multilingual documentation is thus generated directly from the user interface and domain task model, in contrast to the current practice where the initial documentation is produced in one language only and subsequently translated.

The goal of this work package (WP6) is to provide a linguistic specification for the subsequent development of a grammar model which covers the relevant areas of Bulgarian, Czech and Russian in the register of CAD/CAM manuals. This entails the formal description of linguistic structures which are used for the expression of communicative needs in these languages. The tasks of WP6 contribute to the incremental development of the grammar model, ensuring that each stage is self-contained and linguistically interesting and that the final description is robust. In short, this research provides a basis for:

- the development computational grammars of the chosen languages adequate for the generation of well-structured instructions;
- an analysis of the linguistic differences between instructions in different languages.

As stated in the Technical Annexe, page 31 the second task of WP6 (WP 6.2) described in this document is devoted to the specification of linguistic phenomena found in the target texts for the intermediate demonstrator in Bulgarian, Czech and Russian (sample texts in these three languages and their English equivalent are shown in the appendix). These texts present complex procedures for achieving a user's goals in a single style, as illustrated in the following sentences in these languages:

Bulgarian: *Изберете Linetype. След това изберете вида линия на елемента от диалоговия прозорец Select Linetype.*

Czech: *Zvolte Typ čáry. Poté zvolte typ čáry daného elementu z dialogového panelu Výběr typů čar.*

Russian*: Выберите пункт Linetype. Затем выберите тип линии элемента в диалоговом окне Select Linetype.*

English*: Choose Linetype. Then select the element's linetype from the Select Linetype dialogue box.*

The formal specifications of grammatical phenomena in WP 6.2 are based on the analysis of the AGILE CAD/CAM corpus studied in WP3 [CORP], and on the lexical resources and text structures for the initial demonstrator studied in tasks WP 4.1 and WP 5.1. Also, they extend the grammar model reported in the deliverable of task WP 6.1 [SPEC1]. This latter model developed during task 6.1 supported the generation of simple procedures in Bulgarian, Czech and Russian for the initial demonstrator, while the present study is devoted to formal specifications of complex instructional texts written in a single style. WP6.2 activities included the following subtasks:

- selection of grammatical constructions necessary for the expression of complex procedures in Bulgarian, Czech and Russian

- investigation of ways of describing the selected phenomena in terms of Systemic Functional Linguistics
- contrastive analysis of the lexicogrammatical resources of Bulgarian, Czech and Russian (as well as of English and German)
- development of formal specifications for core grammatical functions for Bulgarian, Czech and Russian.

The study whose results are reported in this deliverable has been closely coordinated with activities in the intermediate demonstrator tasks WP 4.2 (Lexical and morphological resources) and WP 5.2 (Text structuring). However, the goals of WP6 are distinct from these activities: while WP4 describes the lexical resources of our corpus, and WP5 studies text structures employed in the register of software manuals, WP6 is devoted to the study of lexicogrammatical features which are general for the grammar of each of the languages, though they are more frequent in this register. As Halliday (1985:372) remarks, "different registers […] do not normally have a special grammar of their own". This view has several implications, which are manifested in two contrasting techniques for the development of a large-scale project in language engineering:

1. creating a minimal *ad hoc* resource set, which is extended for any necessary problem *versus* domain;
2. creating a principled resource set which is aimed at describing the complete grammar and is scaled down to be tuned to a problem domain.

The first approach is corpus-driven, and allows the rapid delivery of a language-processing application. Ide, the development of the grammatical specifications and implementation of lexicogrammatical resources for the initial demonstrator of this project used this technique. In addition, these specifications were heavily based on the functional classification of resources available in the English grammar (with necessary adjustments to realisation statements). This led to the development of the initial demonstrator in just a few months.

In contrast, the work reported on in the present deliverable follows the second approach; thus it has been steered towards the development of a coherent version of grammars for typologically-related languages. Since this approach is driven by options available in the language/grammar as an integral system, the study covers some grammatical phenomena which are absent from the current corpus, but which are relevant for grammatical systems which describe phenomena possible in the register; for example, expression of determination in the nominal group. Recent research in language engineering favours this research trend, for example, see (Henschel, Bateman, 1997) and (Nasr, et al, 1998) for attempts at a proper separation between a general grammatical description of a language and a subgrammar in a problem domain (or a controlled language). This enables us to factor out register-specific data and enhance the extensibility of the grammars developed, thus allowing a possible re-use of these resources in other NLG projects. As far as we are aware, no natural language generators exist for the Eastern European languages being treated in AGILE, which makes our approach advantageous for the community at large.

The drawback of the second approach is that it is time-consuming. For this reason, we have been unable to develop complete grammars of respective languages. So the Eastern European partners involved in grammar development coordinated their activities as much as possible in order to benefit from each other's experience, particularly with respect to their

knowledge of Systemic Functional Linguistic and broadly equivalent descriptions of English, French and German which have been developed already and which have served as examples.

1.1 Basic notions of systemic-functional linguistics

Grammatical descriptions in AGILE are based on principles of systemic-functional linguistics, SFL (within the project a more detailed introduction has been given in the deliverable for WP6.1 [SPEC1]). SFL is a British school of linguistic thought (Halliday 1973) belonging to the tradition of functional approaches to language. The two basic theoretical postulates of SFL according to (Halliday 1978) are:

- the structure of language depends on its function as the communication medium in a social system;
- the fundamental components of language structures are meanings.

In addition to these postulates SFL provides a formal descriptive tool called Systemic Functional Grammar, SFG (Bateman 1992). SFG is functional in that it acknowledges three main "metafunctions" around which languages are organised: the ideational, the interpersonal and the textual; SFG is systemic in that the main focus in description is on the grammatical paradigm (or “system”). The ideational metafunction encodes a language's propositional content. Its grammatical dimension is notably reflected in the clause in the system of transitivity, which gives rise to configurations of processes and the participants and circumstances involved, such as Actor, Goal, Addressee, etc. The interpersonal metafunction encodes speakers' roles in an interaction, their attitudes and evaluations. One of the major grammatical reflexes is the clause systems of mood, which distinguish between declarative, interrogative and imperative and account for the differences in syntactic structure that accompany these different moods. The textual metafunction encodes properties of textual organisation, such as global text structure, coherence and cohesion. In the grammar, this is reflected at the clause level, namely in the systems handling thematic structure and information structure, which in turn are reflected by word order, and at the level of the nominal group, namely in the systems handling determination. The following table contains an analysis of functions in a typical sentence from this register distributed over metafunctions:

Textual	Theme-conjunctive	Theme-topical	Rheme	
Interpersonal	Adjunct	Adjunct	Imperative	Complement
Ideational		Locative	Process	Goal
English	*Then*	*in the Data menu*	*select*	*Linestyle.*
Bulgarian	*След това*	*от менюто Data*	*изберете*	*Linestyle.*
Czech	*Poté*	*v menu Data*	*vyberte*	*Linestyle.*
Russian	*Затем*	*в меню Data*	*выберите*	*пункт Linestyle.*

The source representation in AGILE is designed as a domain model, DM [MODL1], which is independent from all target languages and covers such concepts as CHOOSE*, BUTTON*, DIALOG-BOX*, etc, and their configurations. The dedicated text structuring component (developed in the course of WP5) is responsible for mapping a set of DM statements into a set of statements in terms of general communication-oriented categories. This intermediate resource between the language-independent DM stratum and the stratum of natural language phonology is called the Upper Model, UM (Bateman, 1992), which may

be language-dependent, but should be independent from any one problem domain. UM configurations serve as the uppermost semantic specifications for the lexicogrammar, which are realised as grammatical functions in the systemic-choice network of a particular language. Halliday (1985: xix) remarks that there is no clear line between semantics and functional grammar. However, for the purposes of this deliverable, descriptions of grammatical functions are distinguished from their semantic specifications in UM terms; for example, Action and Actee are UM terms, while *typical* functions which realise them in the grammar are Process and Goal, respectively.

1.2 The AGILE approach to describing functional regions of Slavonic languages

The project presents the first ever attempt at a systemic-functional description of Slavonic languages. In doing this we adhered to basic principles of development of large-scale systemic grammars which proved to be successful for English (Matthiessen, 1996), German (Steiner, Teich, in press), Japanese (Matthiessen, Bateman, 1992) and other languages. The grammatical phenomena to be modelled for the intermediate demonstrator cover quite substantial parts of the lexicogrammars of the target languages of AGILE. The following table shows the distribution of functional regions across metafunctions.

<table>
<tr><td colspan="2"></td><td colspan="3">Ideational</td><td>Interpersonal</td><td>Textual</td></tr>
<tr><td colspan="2"></td><td colspan="2">Logical</td><td>Experiential</td><td></td><td></td></tr>
<tr><td colspan="2">Clause</td><td rowspan="6">complexity</td><td></td><td>Transitivity
Circumstance</td><td>Mood, Polarity, Attitude, Modality, Tense</td><td>Theme, Culmination, Conjunction</td></tr>
<tr><td rowspan="5">groups/phrases</td><td>prep</td><td></td><td>Minor transitivity</td><td></td><td></td></tr>
<tr><td>nom</td><td>Meta-actant</td><td>Nominal-type, Epithet, Qualification, Selection</td><td>Person, Attitude</td><td>Determinat-ion</td></tr>
<tr><td>adj.</td><td>Modi-fier</td><td>Quality-type</td><td></td><td></td></tr>
<tr><td>quant</td><td>Modi-fier</td><td>Quantity-type</td><td></td><td></td></tr>
<tr><td>adv.</td><td>Modi-fier</td><td>Circumstantial</td><td>Comment</td><td>Conjunctive</td></tr>
<tr><td colspan="2"></td><td>complexes</td><td colspan="3">simplexes</td></tr>
</table>

As the result of study of our CAD/CAM corpus and other instructional texts in this register, we chose the following regions for closer investigation:

- at the rank of Clause: expression of nuclear and circumstantial transitivity; aspectual and temporal configurations; diathesis, in particular, expression of active, passive and reflexive voice realisations; word order constraints; clause complexity;
- at the rank of Nominal-group: determination and agreement;
- at the rank of Prepositional-phrase: realisation of semantic functions in nuclear or circumstantial transitivity features.

This very wide range of phenomena to be treated in the grammar of complex procedures and our intention to develop a coherent version of grammars across all three languages prevented us from specifying complete grammars of all three languages. Partners agreed on the following distribution of focal points for investigation:

- Bulgarian: mood, tense, and aspect;
- Czech: word-order, determination, and clause-complexity;
- Russian: transitivity (both nuclear and circumstantial), and diathesis.

Results obtained in this research for a particular language have been extended to others; close typological relations between these Slavonic languages help in cross-linguistic comparison of resulted specifications. While the project presents the first attempt to develop wide-coverage systemic-functional grammars for Slavonic languages, this development has been backed by an existing tradition of applying functional principles in Slavonic linguistics, in particular, by available Functional Generative Descriptions for Czech, and Jakobson-influenced functional studies for Russian.

The following academic studies of grammatical phenomena for respective languages have been used in this research:

- for Bulgarian: (Андрейчин et al, 1977) (Бояджиев et al, 1998);
- for Czech: (Daneš, et al, 1987), (Šmilauer, 1972), and various other sources quoted in the text;
- for Russian: (Шведова, 1970), (Золотова, 1988), (Падучева, 1996).

1.3 Corpus investigation principles

Our study of lexicogrammatical phenomena for the intermediate demonstrator is based on the results reported in the deliverable for WP3 [CORP]. A popular CAD/CAM manual (AutoCAD, 1995) was selected as a source for the corpus preparation because of its wide use and availability in English as well as in all the three Slavonic languages. In addition, we linked our analysis to some samples from other instructional texts in the register of software manuals, for example, from the user manual for Microsoft Word, and in other registers, for example, instructions for cooking, car-repair and a washing machine. The set of lexico-grammatical features as designed for corpus coding in [CORP] on the basis of (Hartley, Paris, 1995) is the following:

- **LGF 1: Rank**
  {CL - Clause, **NGR** - Nominal Group, **PGR** - Prepositional group}
- **LGF 2: Process-type**
  {RP - relational**, MNP -** mental**, VEP -** verbal,
  DMP - directed-material**, NMP -** nondirected-material**}**
- **LGF 3: Finiteness**
  {FIN - finite**, NFIN -** nonfinite}
- **LGF 4: Polarity**
  {POS - positive**, NEG -** negative}
- **LGF 5: Modality**
  {NMD - notmodal**, MDZP -** probability**, MDZU -** usuality**,**
  MDLI - inclination**, MDLA -** ability**, MDLO -** obligation}

- **LGF 6: Mood**
  {IMP - imperative**, INT -** interrogative**, IND -** indicative}
- **LGF 7: Voice**
  {VMID - middle**, VEFA -** active**, VEFP -** passive}
- **LGF 8: Agency-specified**
  {AGSU - agency-specified-user**, AGSP -** agency-specified-program**,**
  AGSA - agency-specified-act**, AGN -** agency-notspecified**, AGG -**
  agency- specified-generic}
- **LGF 9: Member-of-clause-complexity**
  {CCS - clause-complexity-simplex**, CCC -** clause-complexity-complex}
- **LGF 10: Clause-dependency**
  {NDC - notdependent-clause**, DCT -** thematized-dependent-clause**,**
  NDCT - notthematized-dependent-clause}
- **LGF 11: Clause-taxis**
  {PRC - paratactic-relation clause**, HRC -** hypotactic-relation clause}
- **LGF 12: Conjunction-type**
  {CTAD - conjunction-type-additive**, CTAL -** conjunction-type-alternative**,**
  CTT - conjunction-type-temporal**, CTS -** conjunction-type-spatial,
  CTM - conjunction-type-manner, **CTCR** - conjunction-type-causal-
  Reason, **CTCP** - conjunction-type-causal-purpose,
  CTCC - conjunction-type-causal-condition}

In addition to these features the discussion of phenomena is to a certain degree based on available academic grammars for the respective Slavonic languages.

1.4 A notation for specification of grammatical resources

Throughout this text we maintain the following conventions in formal specification of the grammatical features and structures of a target language.

Notation element	Example	Comments
functional elements	Actor, Subject	names of constituents as they are used in systems
system names	MOOD	paradigmatic classifications for features
grammatical features	[feature]	features which classify particular functional elements
selection expressions:		
delicacy	[feature-x : feature-y,...]	This captures the type≠ subtype relation between features
simultaneity	[feature-x & feature-y,...]	choice of several features
realisation statements:		Statements which constrain ways in which features are manifested in language

Notation element	Example	Comments
insert	+Subject	insertion of a constituent
conflate	Subject/Actor	conflation of two constituent into one
expand	Mood(Finite)	
order	Subject ^ Finite	Subject is immediately at the left of Finite
partition	Aux-verb...Predicator	Aux-verb is at the left of Predicator with a possible gap
preselect	Subject:nominal-group	Selection of choice at another rank
lexical constraints:		
classify	Process::doing-verb	Constraints on classification of a functional element
inflectify	Noun:::singular	Constraints on inflectional features of a functional element
lexify	Subordinator ! in-order-to	Choosing a lexical item realising a functional element
agreement	Thing ~ Quality (accusative=accusative)	Propagation of features from one functional element to another

For example, an instrumental circumstance can be expressed in Russian in different ways depending on the process type, the instrument itself and the intention of the speaker. Let us define these choices as choices between [agentive-instrument], [manipulating-instrument] and [enabling-instrument], which are respectively realised by a nominal group in the instrumental case, a prepositional phrase with the simple preposition *с* (*with*), and a prepositional phrase with either of the complex prepositions *с помощью* (*by means of*) or *путем* (*using*) (this last choice is specified by other systems). These possibilities are expressed in our formal notation as:

```
INSTRUMENTAL-PROCESS-TYPE
instrumental-process →
   [agentive-instrument] (Instrument : Nominal-group, instr)
   [manipulating-instrument] (Instrument : prepositional-phrase,
                              Minor-process ! s)
   [enabling-instrument] (Instrument : prepositional-phrase, enabling)
```

This notation is easy rendered into the formal language of KPML, where these choices are expressed as follows:

```
(SYSTEM
    :NAME   INSTRUMENTAL-PROCESS-TYPE
    :INPUTS   INSTRUMENTAL-PROCESS
    :OUTPUTS   ((0.4 AGENTIVE-INSTRUMENT
                     (PRESELECT INSTRUMENT NOMINAL-GROUP)
                     (PRESELECT INSTRUMENT INSTR))
```

```
                (0.3 MANIPULATING-INSTRUMENT
                       (PRESELECT INSTRUMENT PREPOSITIONAL-PHRASE)
                       (LEXIFY MINOR-PROCESS S))
                (0.3 ENABLING-INSTRUMENT
                       (PRESELECT INSTRUMENT PREPOSITIONAL-PHRASE)
                       (PRESELECT INSTRUMENT ENABLING)))
    :CHOOSER   INSTRUMENTAL-PROCESS-TYPE-CHOOSER
    :REGION   PPOTHER
    :METAFUNCTION   IDEATIONAL
)
```

In cases where there is a need to express complex input conditions, as often happens in "gates" (i.e. sets of realisation statements triggered by a configuration of choices), we use comma and round brackets for AND conditions and semicolon and curl braces for OR conditions:

```
IZ
(away-from, {zero-dimension; three-dimension}, contact) →
 (Minorprocess ! IZ, Minirange:Genitive)
```

meaning that IZ is chosen, if away-from and contact features and either zero- or three-dimension features are chosen. In KPML this is expressed as:

```
(GATE
 :NAME   IZ
 :INPUTS   (AND AWAY-FROM (OR ZERO-DIMENSION THREE-DIMENSION) CONTACT)
 :OUTPUTS   ((1.0 IZ
     (LEXIFY MINORPROCESS IZ)
     (PRESELECT MINIRANGE GENITIVE)))
 :REGION   PPSPATIOTEMPORAL
 :METAFUNCTION   IDEATIONAL
)
```

2. The chosen phenomena in the Slavonic languages

2.1 Transitivity

2.1.1 Theoretical background

In systemic-functional grammar the clause is considered as the most significant grammatical unit for representing a mental picture of reality (Halliday, 1985). In general, this covers the experiential resources of a simple clause which are summed up in Halliday's tradition as TRANSITIVITY (systems of process type, agency and different types of circumstances). This experiential structure is realised as a speech act, thus relating to interpersonal structures, and as a contextualised linear message, thus relating to textual structures.

Following Halliday's methodology the TRANSITIVITY region specifies three components in the clause configuration:

1. the process itself;
2. participants in the process;
3. circumstances associated with the process.

As commonly assumed for English as well as for other Indo-European languages, Slavonic languages distinguish between four major process types: [material], [mental], [relational] and [verbal], which are structurally realised as configurations of a Process and the following Participants: Actor-Goal for material, Senser-Phenomenon for mental, Attribute-Carrier for relational and Sayer-Verbiage for verbal processes. Other participants typical for material processes are Agent, Beneficiary, Addressee. Below we do not consider options for mental processes with Senser-Phenomenon and verbal processes with Sayer-Verbiage configurations, because they do not occur in our texts.

Lang	Actor	Material Process	Goal
E	*OFFSET*	*creates*	*copies of lines.*
B	*Командата OFFSET*	*създава*	*копия на линиите.*
C	*Příkaz OFFSET*	*vytváří*	*kopie čar.*
R	*Команда OFFSET*	*создает*	*копии линий.*
Lang	Carrier	Relational Process	Attribute
E	*Line segments*	*are*	*separate objects.*
B	*Отсечките*	*са*	*самостоятелни обекти.*
C	*Úsečkové segmenty*	*jsou*	*samostatné objekty.*
R	*Линейные сегменты*	*являются*	*отдельными объектами.*

Circumstances are partly independent from the process type; they include functions Accompaniment, Cause, Manner, Matter, Role, Spacelocative, Spaceextent and others (they are described in the CIRCUMSTANTIAL region of the grammar). In addition, Matthiessen (1996) remarks that not only participants, but "certain types of circumstance are more restricted with respect to process type than others". For example, circumstances of matter are typically associated with verbal processes (*he asked us about the new model*) and also with certain mental and relational clauses; circumstances of degree are restricted to those processes that can be scaled for intensity (*he deeply regrets your decision*).

2.1.2 Description for Russian

2.1.2.1 Types of Processes

Below we list basic classes of Russian verbs found in our texts with respect to their syntactic patterns (the results that are summarised here are reported in LEXN2 in greater detail).

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References [CORP]
являться	be	Domain Nom Range Inst	40
образовать (образовывать)	form	Domain Nom Range Acc	22
состоять	consist of	Domain Nom Range "из" + Gen	42

Figure 1 Russian: Relational processes

Our corpus contains three relation-process verbs listed in the table above. They are universal with respect to a problem domain. The second and third processes may be expressed as conversions of *состоять из – образовывать* (*X состоит из Y-ов vs. Y-и образуют X (X consists of Y-s vs. Y-s form X*).

The majority of verbs are Dispositive-Material-Processes. In effect, there are some creative material processes, such as *нарисовать* (draw) or *создать* (create), but they show no difference in their government patterns from dispositive ones, such as *открыть* (open) or *нажать* (press). Moreover there are border cases, such as *ввести* (enter: enter a text or enter a key). The following table lists Dispositive-Material-Processes from the corpus.

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
ввести (вводить) (ввод)	enter	Actor Actee (Static-spatial) «в»+ loc	69, 115, 153
(задать) задавать (задание)	set	Actor Actee	25
закрыть (закрывать) (закрытие)	exit close	Actor Actee	80, 107
замкнуть (замыкать) (замыкание)	close	Actor Actee	38
записать записывать (запись)	record	Actor Actee (Destination) «в/на» + acc	175, 191
сохранить сохранять	save	Actor Actee	79, 98

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
сохранение		(Destination) «в/на» + acc	
запустить запускать запуск	start	Actor Actee (Means) + Instr	2, 14, 18, 27
нажать нажимать нажатие	press enter choose select click	Actor Actee	11, 17, 19
описать (описывать) (-)	circumscribe	Actor Actee	157, 170
определить (определять) (определение)	specify	Actor Actee	84
опустить (опускать) (-)	put down	Actor Actee	184
отменить (отменять) (отмена)	undo	Actor Actee	16, 134
открыть (открывать) (открытие)	open	Actor Actee	61
поднять (поднимать) (поднятие)	lift up	Actor Actee	186
показать (показывать) (показ?)	display show list	Actor Actee	91, 119, 120
(отредактировать) редактировать (редактирование)	edit	Actor Actee Manner *с помощью* +gen.	172

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
нарисовать рисовать рисование	draw sketch ∅+line	Actor Actee	1, 26, 43, 175 192 187, 189, 201
создать создавать (создание)	create	Actor Actee	60 24, 87
указать (указывать) указание	specify enter	Actor Actee Manner *в форме* +gen	58 154, 169
выбрать (выбирать) (выбор)	choose	Actor Actee (Spatial-locating) «в, на»+loc	85, 125
отобразить отображать/ся (отображение)	display	Actor Actee (Spatial-locating) «в, на»+loc	87 83

Figure 2 Russian: Dispositive-material-processes

Other classes of processes listed in the table below include happening-processes (*появиться, показаться*) and motion-processes (*перейти, вернуться*). In English *switch* (a translation of *перейти* in the corpus) belongs to Dispositive-Material-Processes, while in Russian it is a motion etymologically and follows motion government pattern with source, destination and route slots (like the English verb *transit*).

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
(-) показываться (-)	show	Medium (Spatial-locating or enumeration) «среди»+gen	92
появиться (появляться) (появление)	appear	Medium (Spatial-locating) «в, на»+loc	215
вернуться	return	Actor	56, 96

(возвращаться) (возвращение)		(Destination) «в»+acc	
перейти (переходить) (переход)	switch	Actor (Destination) «в»+acc	53

Figure 3 Russian: Non-directed material processes

Some processes are expressed by nouns, such as *добавление* (adding): *для добавления элементов в мультилинию* [81]. In English this is expressed by the clause *To add more elements to the multiline*, though this clause in Russian can be restated in the English way *чтобы добавить элементы в мультилинию*. The nominalization alters also the government pattern: *добавить элементы* – Acc; *добавление элементов* – Gen.

Additional feature of Directed-Motion-Process should be added to lexical items, in order to represent material dispositive actions accompanied with a spatial locating circumstance representing a destination. For example, *добавить* (add) has a destination expressed by *в*+accusative: *чтобы добавить элементы в стиль* [68] (*to add elements to the style*). The style here is considered as a container, so the destination role is obligatory in this context. The prefix *до-* includes partly this role – a process of filling the container, as for the Russian verbs: ***до**лить воды в ведро* (*add water to a bucket*), ***до**сыпать грибов в корзинку ...* (*add mushrooms to a basket*).

We also have phase processes for verbs *повторять* (repeat), *завершить* (complete) and *продолжить* (continue). Their specific feature is that Actee is a process expressed by an infinitive verb form or a nominal group representing an action. In contrast to the English *to complete the line,* the Russian *закончить рисование линии* (*to complete **drawing** of the line*) requires an explicit action.

2.1.2.2 Types of Participants

Major differences with respect to the English TRANSITIVITY system arise from the fact that Russian is a highly inflectional language, in which grammatical relations are morphologically marked, as for instance reflected in a case government pattern. See the following example:

(1) *Команда OFFSET создает копии линий.*

OFFSET creates copies of lines

(1a) Командой OFFSET создаются копии линий.

Copies of lines are created by OFFSET

The process type in both (1) and the structural configuration is Actor—Process—Goal. In (1), which is in active voice, the Actor (*команда*) is conflated with Subject and carries nominative case and the Goal is conflated with the Direct Object carrying accusative case. In (1a), the passive version of (1), the Goal becomes Subject and thus carries nominative case and the Actor is expressed by Instrumental case.

Typically participants in the Russian clause are expressed by nominal groups that have an obligatory feature CASE. Two types of case functions are distinguished: direct case

functions (Subject realised by the nominative case, and Direct-complement realised by the accusative case) and oblique case functions (others). The choice of the direct case functions is included in the voice system (see the next section on diathesis); the choice of others depends on the process type and on some rank conditions. In general the functions of Russian Case-Nominal-group are as follows:

Case	Basic meaning
Nom	expresses the Subject, prototypically is conflated with the Actor in the active voice
Acc	expresses the Direct-Complement, prototypically is conflated with the Actee in the active voice
Gen	represents meta-actant relations, i.e. relations between nominalised process and its semantic actant (*рисование линии - the drawing of a line*)
Dat	expresses the Addressee;
Instr	expresses the Instrument; the Agent, if the nominative-case is preselected for another nominal-group; in nominal groups — parameters (*дом высотой 50 м — a house of 50 m*)

Realisation of these functions by case preselection in nominal groups also depends on the government pattern of a lexical item chosen to represent the process. In particular, the two most inherent participants of the process, Actor and Actee, depend on a respective lexeme expressing the process. Prototypically Actee is subjected to modification or treatment, while Actor is considered as Agent, a causer of action. However, which participant is considered as an Actor follows the lexical choice, as evidenced in converse pairs: *купить* (buy) — *продать* (sell); this is the lexical side of diathesis.

The opposition in the role Beneficiary between Client (creative) vs. Recipient (dispositive), as it exists in the English grammar is not present in Slavonic. For example, in Russian, both relations are expressed by either the preposition *для* or by the dative case, depending on the government pattern of the verb.

2.1.2.3 Formal specifications of the nuclear transitivity

The functional structure of the clause is described in two main regions of the grammar - RELATIONALTRANSITIVITY and NONRELATIONALTRANSITIVITY. The case of nominal groups for realisation as participants is preselected upon their insertion. It follows choices from the systems of diathesis as well as in the government pattern of verbs, which is read from their lexical item (in the course of calling the lexical verb term resolution inquiry at the clause rank). In this process in Russian we distinguish between three types of Participants: “direct participants” Actor (Agent) and Goal, which are typically realised by nominal groups in the nominative and accusative cases; “indirect participants” Beneficiary and Addressee, which are typically realised by nominal groups in the dative case; “rest”, which is reserved for the expression of those participants with alternations in diathesis. This includes, for example, expression of Agent by a nominal group in the instrumental case, if it is not Subject; expression of Actor by a nominal group in the genitive case, if the Process polarity is negative. In addition, if a participant is expressed by a prepositional phrase (according to its government pattern), a corresponding preselection is made.

```
ACTOR-INSERT
{(MEDIATED, NONMETEOROLOGICAL, {INDICATIVE; JUSSIVE; IMPERATIVE-OTHER});
  (AGENT-INSERTED, MATERIAL)} →
(+ACTOR: NOMINAL-GROUP)

NONOBLIQUE-COMPLEMENTED
{MED-BEN; AG-BEN; NONBENEFACTIVE; NONCLIENCY; MENTAL; RELATIONAL; (MIDDLE
   VERBAL)} →
(ACTEE: ACCUSATIVE)


DEICTIC-SELECTION
{(MATERIAL-PROCESSUAL-AGENTIVE MATERIAL-PROCESSUAL);
  VERBAL-PROCESSUAL-AGENTIVE} →
[DEICTIC-AGENT] (AGENT/DEICTIC)
[NONDEICTIC-AGENT] (AGENT: NOMINAL-GROUP, INSTRUMENTAL)

NOMINAL-GROUP-CASE
nominal-like-groups→
   [nominative]
   [genitive]
   [dative]
   [accusative]
   [instrumental]
   [prepositional]
```

We have three types of phase processes: *повторять* (repeat), *завершить* (complete) and *продолжить* (continue). So we must modify the systems PHASE-TYPE:

```
PHASE
TRANSITIVITY-UNIT →
[NOT-PHASE]
[PHASE] (+PHASE::PHASEVERB) (+PHASEDEPENDENT), (PHASE/VOICE),
   (PHASEDEPENDENT/LEXVERB)


PHASE-TYPE
PHASE →
[BEGINNING]
[REPEATING]
[ONGOING]
[ENDING]

BEGINNING-TYPE
BEGINNING →
[STARTING]
[BEGIN]

ENDING-TYPE
ENDING →
[STOP]
[FINISH]


ONGOING-TYPE
ONGOING →
[KEEP]
[CONTINUE]

PHASEDEPENDENT-TYPE
{BEGINNING; CONTINUE}→
[PHASEINFINITIVE] (PHASEDEPENDENT::INFINITIVE)
```

```
[NOMINAL-GROUP-PHASE] (LEXVERB:NOMINAL-GROUP) (LEXVERB:ACCUSATIVE)
```

We also need an alternation of realisation of Thing at the nominal group rank, since if it expresses a nominalization of a process, its Actee is realised in the genitive case.

We probably need to introduce into the UM a concept of "forming" as a subclass of identity (a source from which the identified is formed). In the Russian UM, ways for expression of identity differ in their government pattern. For example, *образовывать* + Acc, *состоять из* + Gen, which are lexical converses: *линии образуют фигуру* (*lines form a figure*), *фигура состоит из линий* (*a figure is formed from lines*).

2.1.2.4 Types of Circumstances

This subsection includes a description of Russian prepositions and functions of prepositional phrases and a description of Russian adverbial groups.

2.1.2.4.1 Description of prepositional phrases

Below we list the major Russian prepositions and their syntactic patterns, including those absent from the AGILE corpus, but possible in this or similar registers. We consider two types of prepositions - simple prepositions and derivative prepositions. The simple prepositions can be used in their original meaning (mostly locative) and in a metaphorical extension to abstract notions:

расположен между линиями (situated between two lines) - location,
отношение между составляющими (relation between the constituents) - abstract relation.

Thus they are characterized by the simplicity of their form and the polysemy of the relations they express. The derivative prepositions are derived from meaningful world classes and usually express one particular type of relation.

Most simple prepositions are used with one particular case, several – with two cases, two prepositions (*по* and *с*) are used with three cases (Zolotova, 1988). Below we provide the complete list of simple prepositions in Russian according to the case they govern. Also some semantic oppositions and synonymy relations realised by prepositions are listed. Prepositions that are present in our corpus are given in bold and followed by the semantic relations they express (the inventory of relations is taken from the UM) and their English translation equivalent. Translation equivalents given for prepositions not occurring in our texts list only the basic meaning.

2.1.2.4.1.1 Simple prepositions

Prepositions that occur only with the genitive case:

без = accompaniment [without] [187];

для = **client [for]** [88], **purpose [for]** [14, 154, 169], **[so that]** [187];

до = **up to** [81], [101];

из = from *arc-combination—состоящая из дуг* [42], *one of—один из* [110];

из-за = from behind, because of

из-под = from under

от= from *spatial-locating* [24]

против = against

среди = *in the image tile - среди образцов* [92]

у= *spatial-locating* [at] [87]

Prepositions that occur only with the dative case:

к= to

Prepositions that occur only with the accusative case:

про= about

под = under

сквозь = through

через= *spatial-locating* [through] [24]

Prepositions that occur only with the instrumental case:

над =above

перед = before

Prepositions that occur only with the prepositional case:

при = by

Prepositions that occur with two cases:

в +Acc = *spatial-locating* + *motion-process* +*towards* [in] [81, 154]; **[to]** [53, 55].

+ Loc = *spatial-locating* [during] =temporal-locating[16]

= *spatial-locating* + *zero-dimensional* [at] [21]

= *spatial-locating* + *tree-dimensional* [in] [58, 69]

= *spatial-locating* [under] (*Under* Caps - *B* *разделе Caps)* [88]

= *source* [from] [112]

за +Acc = for

+ Instr = behind

между +Gen

+Instr = between

на +Acc

+ Loc = *spatial-locating* + *one-two-dimension* [on] [5];

о +Acc

+ Loc (close to the meaning of the English "about" does not have locative meaning)

Prepositions that occur with three cases:

по + Dat = temporal-locating or = *spatial-locating extent-process*[at] [182] (***At*** *the Record increment prompt - **По** запросу Record increment)* **[around]** [187];

+ Acc = the interval **[through]** [82] and (idiomatic) ***spatial-locating*** (directed to a boundary *по правую руку*).

To add more elements to the multiline (up to 16), repeat steps 4 ***through*** *7.*

Чтобы добавлять новые элементы в мультилинию (до 16) [81], повторяйте шаги с 4 ***по*** *7 [82].*

+ Loc = temporal (*по прошествии, по окончании*)
с + Gen = *spatial-locating* + *motion-process* + *away-from* [from] [15, 189];
+ Acc (spatial-comparative)
+ Instr = accompaniment-process [*with*]

2.1.2.4.2.2 *Derivative prepositions*

Derivative prepositions are derived from adverbs, participles or prepositional phrases themselves. They follow the government feature of the word they are derived from. So derivative prepositions usually cannot be used with different cases. The list of derivative prepositions given below is not complete, though it contains all the prepositions that occur in the corpus.

Derivative adverbial prepositions have only spatial and temporal meanings. Synonyms which differ slightly in their use are listed together. Two pairs of prepositions (*внутрь* vs. *внутри*; *вперед* vs. *впереди*) express the oppositions of directed vs. non-directed process configurations similar to the opposition expressed by the accusative vs. locative cases of simple prepositions, like *в*.

Derivative adverbial prepositions that occur only with the genitive case:

вблизи (nearby)

вдоль (along)

вне (outside)

внутри vs. внутрь (inside)

возле, около, подле (nearby)

вокруг = *spatial-locating* [about] [170], кругом

вперед vs. впереди (in front of)

мимо (beside)

напротив (opposite to)

поверх (on top of)

позади, сзади (from the rear)

поперек (across)

после (after)

посреди (in the middle of)

прежде (before)

сбоку (at the side of)

сверху (from above)

свыше (over) **...**

Derivative prepositions with the genitive case express a static spatial-locative or a temporal-locative semantic relation.

Derivative adverbial prepositions that occur only with the dative case. Two of them have directive-locative meaning (the goal is a moving object):

вслед (following)

навстречу (towards),

Others have abstract, logical meanings. They realise, with different connotations, an opposition *contrary to* vs. *according to*:

вопреки, наперекор (in spite of)

согласно (Gen, Inst), сообразно (Inst, Loc), соответственно, соразмерно (Inst, Loc) (accordingly)

Prepositions that are derived from nouns occur only with the genitive case as a universal relation between two nouns in Russian (with the exception of ***по отношению к*** occurring with the dative case following the government pattern of the preposition *к*). They all have abstract meanings:

ввиду = in view of

в качестве = as [103]

в меру = as far as

в области = in the region of

в отношении = with respect to

по отношению к = in relation to [57]

в продолжение = in the continuation of

в силу = by virtue of

в смысле = in the sense

вроде, наподобие = similar to

вследствие = because of

в течение = during

в целях = in order to

за исключением = with the exception of

насчет = concerning

по линии = along

по мере = as long as

по поводу = on the occasion of

по причине = because of

по случаю = on the occasion of

по части = as regards

со стороны = on the part of

с помощью, посредством, путем **= [by means of]** [173]

Prepositions that are derived from verbs occur with different cases mostly following a verbal government pattern. Often, prepositions of this type provide oppositions:

благодаря + Dat vs. несмотря на + Acc (due to vs. in spite of)

включая + Acc [101] vs. исключая + Acc (including-comitative vs. excluding-comitative)

кончая + Instr vs. начиная + Instr (ending with vs. beginning with)

спустя + Acc (after, temporal)

2.1.2.4.2 Formal specifications for prepositional phrases

Specifications are based on the UM roles and the English grammar model (PPOTHER and PPSPATIOTEMPORAL regions). Prepositions are treated as realisations of the following type of semantic relations from the UM **process : relational-process : two-place-relation : circumstantial**. They are represented as relations between two entities: a main process or an object (Domain) and an object (Minirange). Following the SFG terminology, they are seen as minor-processes having several possibilities for realisation depending on semantic properties of the minor-process itself and of its Minirange. The semantic properties of the **circumstantial** are further classified in the ideational metafunction following the UM as: **spatio-temporal, causal-relation, ordering-relation, subject-matter, generalized-means, comparison,** and **accompaniment.**

Some systems of the English grammar are relevant for the generation of Russian prepositional phrases, and in these cases changes are minimal. Others presuppose splitting a system or creating a new one. The most diverse network describes locative prepositions, since the locative semantics is widely represented in our texts. Temporal systems are rarely found in our texts, though they are often described by similar oppositions, for example, extent vs. location. Most prepositions are used with one particular case (see the list above), so the selection of a minor process is directly followed by a case inflection constraint on its Minirange, as in ACCOMPANIMENT-PROCESS-TYPE:

```
[excluding-comitative]  (minorprocess ! bez, minirange:genitive)
```

But some simple prepositions are used with several cases, expressing different relations. This is the case of a double selection: lexicalization of minor-process and the choice of the case of its Minirange. We may specify the case for each context of a preposition. However, in this research we tried to investigate oppositions that are available in the Russian semantics, in particular, such oppositions as directed vs. non-directed realised by the accusative and prepositional cases respectively, destination vs. location realised by accusative and genitive cases (cf. above-mentioned *внутрь—внутри* realizing the same opposition by lexical means). In our treatment we tried to keep as close to the English oppositions as possible, whenever they are present in the Russian lexicogrammar.

2.1.2.4.2.1 Spatiotemporal prepositions

A top level distinction within spatiotemporal relations is between **spatial-process** and **temporal process**. In this section we consider only spatial processes (the only temporal meaning that occurs in our corpus does not fall within the scope of the intermediate demonstrator). Another top level distinction within spatio-temporal relations is that between **extent-process** (picks up a segment of space-time) and **location-process** (picks up a point

of location). Subtypes of **extent-process** are **absolute-extent-process** (*to go for three weeks; to extend for two meters*) and **relative-extent-process.** The gate FOR of the English lexicogrammar serves the realisation of the former subtype together with the relations **client** and **purposive-process**. In Russian these relations are realised by different prepositions: *на* - *ехать на три недели; увеличить на два метра* - for the **absolute-extent-process** and *для* for the **client** and **purposive-process** features. So instead of a single English gate FOR we introduce two Russian gates: DLJA-GOAL with the input: **client** or **purposive-process,** and ABSOLUTE-EXTENT-PROCESS. The feature **relative-extent-process** (as it exists in the English lexicogrammar) is not present in our texts.

There is a construction which is missing from the English grammar, generation of the proposition *through* (*через* in Russian). English has a more delicate system depending on the dimensionality of Minirange - *across the river* (**through the river*), but *through the point.* Russian lacks this distinction - *через реку, через точку.* The EXTENT-PARALLELISM system lexifies the minor-process as *через* and preselects Acc for the Minirange:

```
[EXTENT-PARALLELISM] (minor-process ! Cherez, Minirange:Acc)
```

The two top level distinctions for the **location-processes** in the English model are choices between a type of the minor-process (**motion** vs. **rest**) and a status of the Minirange in relation to the minor-process, whether it is a PLACE itself, or an orientation axis for the PLACE, (**nonorientation-axis** vs. **orientation-axis**, respectively). **Orientation-axis** is not relevant for the **motion** processes. According to orientation type (**vertical-orientation** vs. **horizontal-orientation** and their subtypes) the **rest** minor-processes are realised by prepositions *facing, above, between,* and so on. This type is not used in our texts. In Russian the same preposition is used in both non-orientation and orientation meanings: *на доме - на дом, под домом - под дом, побежать за дом (за домом),* but **перед дом (к дому) *между* –only rest.

The functions of English prepositional groups follow the classification of functions of ZERO-DIMENSIONAL (realised by *at*) vs. ONE-TWO-DIMENSIONAL (realised by *on*) vs. THREE-DIMENSIONAL (realised by *in*). The lexical stock of prepositions available in Russian distinguishes only two functions: ONE-TWO-DIMENSIONAL (realised by *на*) vs. ZERO-THREE-DIMENSIONAL (realised by *в*). The model of dimensions also depends on the semantic features of a process, which define a metaphorical interpretation for an object's dimensionality in language: *они уставились* ***на*** *пещеру/****на*** *лицо* (two-dimensional), *they stared* ***at*** *the cave/****at*** *the face* (zero-dimensional), - but *они вглядывались* ***в*** *пещеру/****в*** *лицо they peered* ***into*** *the cave/****into*** *the face* (three-dimensional). The first case implies (at least in Russian) that the view is directed at the surface of an object, while in the second case the view penetrates into this object (either the cave or person's psyche).

Directness. In English, circumstances of directed location are restricted to those processes that can be interpreted in terms of a *vector (she kicked the ball into the corner; they gazed into the cave*) (Matthiessen, 1996). **Nonorientation-axis** is chosen for treatment of the Minirange as a destination or a source of the process. Here realisation of the minor-process by a preposition depends only on the type of the Minirange. In the English model, three types are distinguished: **zero-dimension, one-two-dimensions, three-dimensions.** In the latter case the Minirange is considered as a sort of a container: **contained** vs. **not-contained.** Two standard means in Russian are locative prepositions *на* (**one-two-dimensions**) and *в* (**zero-** or **three-dimension**) depending only on the dimensionality of the PLACE.

The spatial functions of Russian cases are described by the following oppositions: Acc vs. Loc or Gen. Spatial prepositional phrases with Acc realise the directed motion:

(1) towards a location - *в дом, на площадь*

(2) towards a border. This meaning appears, when the main process does not express directed movement - *с 4 по 7*; *(сидеть) по правую руку*. In English this interval is expressed by a "through" process: *repeat steps 4 through 7* vs. *повторяйте шаги с 4 по 7 [82].* Another type of interval is *от — до* (from — to), in which both limits are realised by the genitive case. Interval meanings appear in our texts, but are not chosen for the intermediate demonstrator, so they are not treated below.

(3) towards a location relative to an object, for example, *за дом, на дом, сквозь дом, через дом* ...

Basically, the genitive case in locative meaning corresponds to the source, and the prepositional case to the location. Use of the accusative case for movement directed to an object implies that it is reached (the contact feature is chosen). If the object is not reached in this movement, another construction is used (the dative case: *к дому — от дома*). These factors have priority over dimensionality of the object.

The feature main-motion-process is inserted in cases when predicate of the clause is expressed by a motion process. The following is a formal specification of basic Russian prepositions.

```
MOTION-PROCESS-TYPE
(motion-process, main-motion-process) →
   [away-from] (Minirange : gen)
   [towards] (Minirange : acc)

CONTACT-TYPE
(motion-process, main-motion-process) →
   [contact]
   [non-contact]

V-STAT
{containment-implicit; strong-inclusive; in-extent}→
   (Minorprocess ! v)
   (Minirange : prep)

NA-STAT
{weak-inclusive; (one-two-dimensions, rest-process)}→
   (Minorprocess ! na)
   (Minirange : prep)

V-MOTION
({zero-dimension; three-dimension}, towards, nonorientation-axis,
   contact)→
   (Minorprocess ! v)
   (Minirange : acc)

NA-MOTION
(one-two-dimensions, towards, contact) →
   (Minorprocess ! na)
   (Minirange : acc)
```

```
IZ
(away-from, {zero-dimension; three-dimension}, contact) →
   (Minorprocess ! IZ)

S
(away-from, one-two-dimensions, contact) →
   (Minorprocess ! s)

K-MOTION
(TOWARDS NONORIENTATION-AXIS, NON-CONTACT)→
   (Minorprocess ! k)
   (Minirange : dat)

OT
(away-from, non-contact) →
   (Minorprocess ! ot)
```

The existing English model lacks the prepositions *вокруг* and *среди* (*around* and *among*).In the intermediate demonstrator we have such usage as *среди названий* (literally *among names*) rendered in English as *in a list of names*. Both prepositions are considered subtypes of ORIENTATION "a specification relative to some dimension or orientation axis, e.g., with respect to a vertical plane, in front vs. behind…" (Bateman, et al, 1990). So we add a three-dimensional orientation.

```
ORIENTATION-TYPE
(orientation-axis, rest-process) →
   [vertical-orientation]
   [horizontal-orientation]
   [three-dimensional-orientation] (Minirange : gen)

ANOTHER-ORIENTATION-TYPE
another-orientation →
   [vokrug]
   [sredi]
   [unspecified]
```

2.1.2.4.2.2 Other prepositions

In contrast to the realisation in English of MANIPULATING-INSTRUMENT and INCLUDING-COMITATIVE by nominal groups with markers, Russian uses either nominal groups in the instrumental case, or prepositions (*с* with the instrumental case or *с помощью* with the genitive case):

```
ACCOMPANIMENT-PROCESS-TYPE
accompaniment-process →
   [including-comitative] (Minorprocess ! s, Minirange : instr)
   [excluding-comitative] (Minorprocess ! bez, Minirange : gen)
   [cumulative-additive] (Minorprocess ! vmeste-s, Minirange : instr)
   [alternative-additive] (Minorprocess ! vmesto, Minirange : gen)

INSTRUMENTAL-PROCESS-TYPE
instrumental-process →
   [agentive-instrument] (Instrument : Nominal-group, instr)
   [manipulating-instrument] (Instrument : prepositional-phrase,
                              Minor-process ! s)
   [enabling-instrument] (Instrument : prepositional-phrase, enabling)
```

```
ENABLING-TYPE
enabling →
   [explicit-enabling] (Minorprocess ! s-pomoschju, Minirange : gen)
   [nonexplicit-enabling] (Minorprocess ! putjom, Minirange : gen)

DLJA-GOAL
{client; purposive-process}→
   (Minorprocess ! dlja, Minirange : gen)
```

2.1.2.4.3 Description of adverbial groups

The adverbs in the Russian corpus fall into two groups: some of them are circumstances in a clause - *first/сначала*, *then/затем*, *now/теперь*. Others function as a modifier of a predicate - *newly/вновь, already/уже, again/снова*. Broadly speaking they are all temporal circumstances, though they do not denote time (as in *It snows in winter*), rather temporal conditions for an event presented in the clause. For this reason their choice is closely related to choice of the aspect of a predicate. Also the adverb *теперь* (*now*) is related not to ideational characteristics of the event, but to the mode of its presentation as deployment in time; so it is related to discourse characteristics. Adverbial groups in a clause are thematic in Halliday's sense; they serve as a point of departure, more precisely a connector - an element that relates two propositions in the text.

Corpus References	Russian	English	Type
43	сначала	first	step, connector
125, 127	затем	and	Sequence, connector
74, 77, 94, 149, 164	затем	then	Sequence, connector
97, 129	теперь	now	(discourse, sequence), connector
102	вновь	newly	Temporal, manner
188	снова	again	Temporal, manner
193	уже	already	Temporal, manner
190	каждый раз, когда	at any time to (...) (periodically)	Simultaneity, connector

Figure 4 Russian: Adverbial groups

The positions of adverbial groups in Russian sentences does not significantly differ from the word order in an English sentence. The only case of deviation in word order is the following:

Click the button *again* to resume drawing from the new cursor position.

Снова нажмите эту кнопку [188], чтобы продолжить рисование с новой позиции курсора [189].

2.1.3 Description for Bulgarian

2.1.3.1 Types of Processes

The system of Transitivity belongs to the experiential (ideational) metafunction; it is the system which allows various constructions of what is happening, mainly through choice of alternative types of processes and participants.

As in other Slavonic languages and English, the following types of processes can be employed to this end: material, relational, mental, verbal, etc. In the Bulgarian corpus, which is inevitably constrained by the instructional nature of the text being modelled, two types of processes occur predominantly, i.e., relational and dispositive-material. Here we list the basic classes of Bulgarian verbs found in our corpus; the results summarily presented here are reported in greater detail in LEXN2. We begin our description with relational processes:

There are very few relational process verbs in the Bulgarian corpus, for example, *съм* («be»), e.g.,*броят на елементите е нечетен, ако моливът е поставен, стойността е по-голяма,* etc. and the verb *съдържам (*«contain»), e.g. *което може да съдържа*, etc. Unlike the Russian corpus, in the Bulgarian one there is one occurrence of a mental process verb for perception, e.g., *виждам* («see») as in *ако искате да видите основния цвят* («if you want to see the main colour»). However, just as in the other two languages, this corpus includes mostly verbs associated with dispositive-material processes and these are listed in the following table.

Wordbasis	**English Gloss**	**Corpus Occurrences frame**	**Corpus References**
Взема	take	Actor Actee	154
Въведа	enter	Actor Actee Spatial-locating «в»	13, 16, 31 … 66 …
Добавя	add	Actor Actee Spatial-locating «в»	68 78
Държа	hold	Actor Actee	167
Задам	specify	Actor Actee	8, 9, 10, 27, 28, 151 …
Запиша	save record	Actor Actee	76, 166
Затворя	close	Actor Actee	32, 121
Избера	choose	Actor Actee Actor Spatial-locating «в, на»	65, 67, 69 … 5, 7, 24, 26, 71, 82 …
Използвам	use	Actor Actee	3, 22, 36, 55 …
Изтрия	undo erase	Actor Actee	184 15
Местя	move	Actor Actee Spatial-locating «по»	155

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
Натисна	press	Actor Actee Spatial-locating «на, върху»	11, 29, 51 …
Начертая	draw	Actor Actee	34, 150, 163, 182 …
Определя	specify enter	Actor Actee	132, 135
Отворя	open	Actor Actee	58
Отменя	undo cancel	Actor Actee	188
Подновя	resume	Actor Actee	169
Получа	get list	Actor Actee=list+Proper-ty_Ascription «на»	104, 105
Посоча	indicate	Actor Actee	134
Поставям	put down	Actor Actee	148
Превключа	switch	Actor Agentive «на»	44
Премахна	undo	Actor Actee	116
Преместя	move	Actor Actee Spatial-locating «в»	181
Прибавям	add	Actor Actee Agentive «към»	64
Променя	change	Actor Actee	110
Редактирам	edit	Actor Actee Instr «с»	54, 137
Свържа	connect	Actor Actee Agentive «с»	14

Wordbasis	English Gloss	Corpus Occurrences frame	Corpus References
Сложа	put	Actor Actee	164
Стартирам	start	Actor Actee Actor Actee Instr «с»	2, 18, 21, 95 125
Създам	create	Actor Actee	53, 136
Съхраня	record save	Actor Actee Spatial-locating «в»	51
Трия	erase	Actor Actee	192
Чертая	draw	Actor Actee	156, 161, 165

Figure 5 Bulgarian: Process types

As in Russian, other classes of verbs include happening-processes (*показвам се, появявам се*) and motion-processes (*върна се, изляза*). Their corpus occurrence frame in Bulgarian is similar to the specifications in Russian. These are all non-directed material processes. However, not every conclusion for Russian is directly applicable to Bulgarian. One particular difference from the Russian data, for example, is the relatively modest degree of nominalization in purpose-related constructions. Compare, for example, the Russian *для добавления элементов в мультилинию* («for the addition of elements in the multiline») with the Bulgarian equivalent *за да добавите още елементи към мултилинията* («in order to add elements to the multiline»), both of which equate to to the English *To add more elements to the multiline*.

The Bulgarian corpus also offers examples of occurrences of phase processes, e.g., *завърша* («complete»), *повторя* («repeat»), *продължа* («continue»), *приключа* («finish»).

2.1.3.2 Types of Participants

Bulgarian offers an interesting combination of features, some of which bring it closer to English than to the other Slavonic languages. Unlike most of them, Bulgarian has already lost the rich system of noun and adjective inflections which are so characteristic of languages with the grammatical category of case. In modern Bulgarian, case marking remains only in some pronouns such as *кой* (who_Nom), *кого* (who_Acc), *кому* (who_Dat). However, the current trend is for case marking to disappear rapidly from even those few constructions in which it used to occur not so long as a couple of generations ago. Thus, the third form of the above (*кому*) is almost entirely obsolete nowadays—the three-tier paradigm has collapsed into two and is undergoing further neutralization into a singular case-unmarked form (*кой*). It is worth emphasizing here that the majority of word classes do not have forms distinguishing among different cases.

The above analysis shows that in Bulgarian, unlike Russian and Czech, the category of case is not used to distinguish among participants in the clause. They are instead expressed and recognized by means of word order variation, agreement patterns, and suprasegmental information in spoken discourse. Thus, clauses may contain the same participants, although may be perceived as performing different roles, depending on the preferred word order (Bulgarian has a canonical Subject-Verb-Object word order) and the presence or absence of agreement in number and gender of the two noun phrases.

(1) *децата повикаха майката* (children_Def call_PastTense_Pl mother_Def)
(2) *майката повикаха децата* (mother_Def call_PastTense_Pl children_Def)

For example, both (1) and (2) refer to the same situation in terms of the ideational metafunction, i.e., the children called the mother, irrespective of word order variation because of the strength of the Subject-Verb agreement cue. However, if the requirement for agreement is fulfilled by both noun phrases in the same clause, then the preferred (canonical) word order governs the choice of interpretation and encoding.

(3) *детето повика майката* (child_Def call_PastTense_Sg mother_Def)
(4) *майката повика детето* (mother_Def call_PastTense-Sg child_Def)

Examples (3) and (4) illustrate the relevance of word order in case of neutralization of Subject-Verb agreement. The difference in both (1-2) and in (3-4) is structural in the sense of reversal of the order of the two noun phrases, but the effect in the second pair is to produce a different meaning realisation. The clause in (3) has the meaning of «the child called the mother» whereas in (4), the meaning is «the mother called the child».

2.1.3.3 Formal specifications of the nuclear transitivity

The lack of overt case system in Bulgarian language makes the organization of formal representation much closer to the English grammar in comparison with Czech and Russian formal specifications dealing with Transitivity. There are some differences, for example the group of systems related to PHASE VERBS could be changed as follows:

```
PHASE
TRANSITIVITY-UNIT →
[NOT-PHASE]
[PHASE] (+PHASE::PHASEVERB) (+PHASEDEPENDENT), (PHASE/VOICE),
   (PHASEDEPENDENT/LEXVERB)

PHASE-TYPE
PHASE →
[BEGINNING]
[REPEATING]
[ONGOING]
[ENDING]

BEGINNING-TYPE
BEGINNING →
[STARTING]
[BEGIN]
```

```
ENDING-TYPE
ENDING →
[STOP]
[FINISH]

PHASEDEPENDENT-TYPE
{BEGIN; ONGOING; ENDING}→
[PHASEINFINITIVE] (PHASEDEPENDENT:::PRESENT-FORM, +DAPHASE, DAPHASE!DA,
   DAPHASE^PHASEDEPENDENT)
[NOMINAL-GROUP-PHASE] (LEXVERB:NOMINAL-GROUP)
```

2.1.3.4 Types of Circumstances

2.1.3.4.1 Description of prepositional phrases

Prepositions belong to the closed class of words and express relations among phenomena which are referred to by open class words such as nouns, verbs, adjectives, etc. In Bulgarian, they may be used to link up almost any combination of parts of speech (from noun-noun to numeral-adverb to adjective-pronoun) but they primarily express relations among nouns and other parts of speech functioning as nouns. That is why prepositions may be said to express relations among phenomena named by nouns, pronouns, etc., which is in essence the semantic content of the morphological category of case.

The classification of prepositions in terms of semantics is a particularly difficult enterprise, as in analytic languages such as Bulgarian (and English) most prepositions are polyfunctional. This kind of taxonomical effort could be carried out on the basis of semantic relations expressed by traditional case categories (Accusative, Dative, etc.) perhaps, but it does not reflect the linguistic competence and usage of contemporary Bulgarian speakers. Moreover, presumably all prepositions had a spatial-locative meaning originally, which has now been lost, except for those under (a) below.

One possible classification of Bulgarian prepositions is based on the following broad semantic spheres (The examples of prepositional phrases with their English equivalents are taken from the AGILE corpus):

(a) spatial relations: *в, върху, до, зад, из, между, на, над, под, пред, през, при, след, у, от,* etc.

Можете да започнете нова линия ***от*** *крайната точка - You can start a new line* ***at*** *the endpoint*

да излезете ***от*** *диалоговия прозорец- exit the dialog box.*

до *подсказващото съобщение* ***- at*** *the prompt*

на *функционалния ред -* ***on*** *the toolbar*

по *линията -* ***along*** *the line*

в *диалоговия прозорец -* ***in*** *the dialog box*

след *подсказващото съобщение -* ***at*** *the prompt*

по *екрана -* ***around*** *the screen*

да се върнете ***на*** *подсказващото съобщение - return* ***to*** *the prompt*

(b) temporal relations: *в, до, на, от, по, преди, през, след, сред,* etc.

(c) manner relations: *без, освен, от, по, с, чрез,* etc.

(d) causal relations: *за, от, по, поради,* etc.

(e) goal-oriented relations: *до, за, заради, към*, etc.

(f) means-related, instrumental relations: *с (със), на, чрез, посредством,* etc.

редактирате ***с*** *PEDIT - edit* ***with*** *PEDIT*

(g) relations of reference to opinion, compliance, etc.: *според, по*

(h) relations of reference to subject-matter: *за, по, например,* etc.

за *всеки край на мултилинията -* ***for*** *each end of the multiline*

(i) relations of possession: *на*

точка ***на*** *полилинията - point* ***of*** *the polyline*

Note that this phrase may be considered in some contexts to express a spatial relation as in (a) above.

(j) relations of origin/ part-whole relations: *от*

един ***от*** *следните методи - one* ***of*** *the following methods*

от *плаващото меню -* ***from*** *the flyout*

двойки ***от*** *елементи - pairs* ***of*** *elements*

(k) relations of quantity: *на, с*

(l) relations of transcending a boundary, etc.: *свръх*

As the above examples illustrate, most prepositions used in the corpus fall into a limited number of semantic spheres, i.e., spatial, instrumental, and origin /part-whole relations.

2.1.3.4.2 Formal specifications for prepositional phrases

The formal specifications of prepositions used in Bulgarian language seem to be much near to the English ones than to those of Russian or Czech. The main reason is the lack of cases in Bulgarian. On the other hand, some changes in grammatical resources are needed to represent the proper relations Domain – Minirange in particular areas of the Upper Model.

The most widely used relations in the AGILE corpus expressed by means of prepositions are the spatio-temporal relations. For Bulgarian, the basic spatio-temporal prepositions may be specified as follows:

```
MOTION-PROCESS-TYPE
(motion-process) →
   [away-from]
   [towards]

V-STAT
{containment-implicit; strong-inclusive; in-extent}→
   (Minorprocess ! v)

NA-STAT
{weak-inclusive; (one-two-dimensions, rest-process)}→
   (Minorprocess ! na)

V-MOTION
```

```
({zero-dimension; three-dimension}, towards, nonorientation-axis)→
   (Minorprocess ! v)

NA-MOTION
(one-two-dimensions, towards) →
   (Minorprocess ! na)

OT
(away-from, nonorientation-axis) →
   (Minorprocess ! ot)

KAM
(towards, one-two-dimensions, nonorientation-axis)→
   (Minorprocess ! kam)
```

2.1.3.4.3 Description of adverbial groups

Bulgarian adverbs have only one wordform and do not agree with any other word in the sentence. Some adverbs have comparative and superlative degree. Depending on their meaning the adverbs are divided into groups:

- manner
- quantity and degree
- temporal
- spatial
- reason and purpose
- modal

Adverbial groups with an adverb as head, usually have the syntactic function of circumstantial elements. In Bulgarian, the circumstantial elements are expressed mostly by a nominal group or by an adverbial group. The following types of circumstantial elements can have the structure of adverbial groups: temporal, spatial, manner, quantity and degree, modal.

In the sentence, spatial and temporal adverbs normally occupy initial position, and the manner and quantity adverbs are closely related to the predicated, therefore they are very often placed before or after it.

Adverbial groups in the Bulgarian corpus for the intermediate demonstrator are very limited in terms of number and types. We found only 15 instances of adverbs, which are as follows:

Corpus References	Bulgarian	English	Type
70	След това	then	temporal
73	след това	then	temporal
98	сега	now	temporal
102.2	незадължително	optional (not obligatory)	modal
103	включително	including	manner
128	сега	now	temporal
162	след това	then	temporal

Corpus References	Bulgarian	English	Type
180	долу	down	spatial
185.4	отново	again	manner
191.5	периодично	at any time (periodically)	manner
194	вече	already	temporal
214.1	независимо	independently	manner
217	сам	alone (by itself)	manner
227	след като	after	temporal
227	вече	already	temporal
231	след като	After	temporal

By definition an adverbial group has an adverb as head. All adverbial groubs consist of one adverb, which is not accompanied by modifying elements. Most adverbs are of the temporal and manner type, although spatial adverbs can appear as well. The order of the adverbial groups in the Bulgarian sentence does not differ significantly from the word order in an English sentence.

2.1.4 Description for Czech

2.1.4.1 Types of Processes

2.1.4.1.1 Evidence from the corpus

For a detailed study of the data, see LEXN2 (WP4 deliverable 2). In the analysis of transitivity in the Czech corpus, we employed the usual SFG classification of processes as follows:

- material: non-directed, directed (dispositive, creative)
- mental
- relational: intensive, circumstantial, possessive
- behavioural
- verbal
- existential

For each type of process SFG specifies what Participants accompany it. We have identified occurrences of all the major types of processes in our Czech AGILE corpus, except existential processes.

2.1.4.1.1.1 Relational processes

There is a number of occurrences of the verbs *být (be)* and *mít (have)* expressing relations, and also one occurrence of the verb *navazovat (continue)* which we classified as relational. All the three essential types of relations are represented in the corpus. We provide examples of each represented subtype below.

(1) *intensive:* 'x *is* a'

(a) *KŘIVKA kreslí polyúsečku složenou ze segmentů úseček a oblouků, jejíž forma je jeden objekt.* [222-23]
PLINE draws a polyline consisting of line segments and arcs whose form is one object.

(b) *Jestliže je Krok menší, pak má přírůstek záznamu prioritu.* [252-253]

(2) *circumstantial:* 'x *is at* a'

(a) *Nakreslíte rovný segment křivky, na který navazuje obloukový segment* [45]
You will draw a straight line segment continued by an arc segment

(b) *Výsledná křivka je umístěna podél středu kde byla křivka.* [75-76]
The resulting line is placed along the center where was the line.

(c) *Kreslení od ruky začne, ať je kursor kdekoli.* [240-241]
Sketching begins no matter where the cursor is.

(d) *Jestliže je pero dole, …* [234]
If the pen is down …

(3) *possessive:* 'x *has* a'

(a) *Jestliže máme sedm elementů, …* [117]
If we have seven elements

(b) *Pod* Popis *zadejte popis, který může mít až 255 znaků včetně mezer.* [135]
Under Descriprion enter a description which may have upt to 255 characters including spaces.

2.1.4.1.1.2 Behavioural processes

We have found one verb in the Czech AGILE corpus which we believe should be classified as expressing a behavioural, near mental, process, namely *přehlížet (ignore)*.

(4) Nastavení kroku *přehlíží nastavení kroku.* [250]
Specify step *ignores the step setting.*

2.1.4.1.1.3 Material

We have found one verb which should be classified as expressing a **material non-directed process (happening)**, namely *objevit se (appear)*.

(5)

(a) *Jestliže vytvoříte element s negativním offsetem, objeví se pod originálem v obrazovém políčku.* [103-104]
If you create an element with a negative offset, it appears under the original in the picture window.

(b) *Tato barva se neobjeví v obrazovém políčku v dialogovém panelu* Styly multičár. [126]
This color will not appear in the picture window in the dialog box Multiline Styles.

The vast majority of verbs occurring in our Czech corpus can be classified as realizing directed material processes. All the material processes combine with a Goal, and are therefore directional.

According to the SFG view, **directed material processes** have two Participants, **Actor** and **Goal**. Furthermore, it is possible to distinguish between **dispositive** (doing to) and **creative** (bringing about) processes. Most of the verbs in our corpus belong to the dispositive type, but we also found verb occurrences that we would classify as creative. Some verbs are borderline cases, because it is difficult to decide whether the Goal has existed before the action or is the result of the action. Nevertheless, the table in Figure 6 lists the occurrences of verbs and their nominalizations in our corpus that we classified as creative, not requiring any other complementations than Actor and Goal. The table in Figure 7 lists the dispositive ones, which do not require any other complementations than Actor and Goal. Besides the verbs summarized in the tables in Figure 6 and Figure 7, we also encountered verbs which have complementations that appear to go beyond the classification provided in SFG. First of all, there are a number of verbs which semantically require a Circumstantial Spatial-locative complement. For example, the verb *zobrazit (display)* in Czech requires a Location where the Goal should be displayed. A number of the Czech verbs require a Direction towards which the process is oriented, e.g. *přidat (add), zapsat (record)*. These are summarized in the table in Figure 8.

The usage of the verbs in table Figure 8 can be illustrated using the following examples which are simplifying modifications of sentences occurring in the corpus.

(6) *Location-where*

(a) *Tento příkaz zobrazuje čáry ve vrcholech multičáry*
This command displays lines in the vertices of a multiline.

(7) Direction

(a) *Tento příkaz přidává elementy ke stylu*
This command adds elements to a style.

(b) *Tento příkaz zapisuje elementy do databáze.*
This command record elements into the database..

Moreover, there are several verbs in the corpus, which besides Actor and Goal require semantically a complement that appears to differ from a normal Circumstantial Spatial-locative. The "surface" difference is that such complements have very restricted ways of realisation, e.g. only particular prepositional phrases. not adverbs. But more importantly, they seem to express conceptually different entities from Spatial-locations. We distinguish two such complements and call them *Effect* and *Origin*. The verbs requiring such

complementations in Czech and occurring in the Czech AGILE corpus are summarized in the table in Figure 8 Czech: Directed material processes requiring a Spatial-locative complement.

can be illustrated using the following examples, which are simplifying modifications of sentences occurring in the corpus.

(8) Effect

(a) *Tento příkaz spojí čáry do párů*
This command conjoins lines into pairs

(b) *Tento příkaz vytváří obdélník z úseček*
This command creates a rectangle out of lines

(c) *Tento příkaz mění režim na Zarovnávání*
This command changes the mode to Justification.

(9) Origin

(a) *Tento příkaz přepíná z režimu Krok do režimu Orto.*
This command switches from the Step mode to the Orto mode.

(b) *Tento příkaz vybírá z dostupných barev barvu elementu.*
This command selects element's colour out of the available colours.

(c) *Nyní z nabízených možností zvolte typ zarovnání čar.*
Now select the type of line alignment out of the offered possibilities.

Next, we shall rephrase these observations in terms of valency frames in the framework of FGD (Sgall et al. 1986). The gain of doing so will become apparent when relating the issues raised by the observations to the structure of the Upper Model (UM).

Wordbasis	Gloss	Theoretical valency frame	Realisation of Participants in the Corpus	Corpus References
definovat	define	Act Goal	Goal-acc	47,87
dokončit dokončení	finish	Act Goal	Goal-gen	281
kreslit kreslení	draw	Act Goal	Goal-acc Goal-gen	1,27,59,63 22,232
nakreslit nakreslení	draw-infinite	Act Goal	Goal-acc Goal-gen	44,88 50,147,179
nastavit nastavení	set	Act Goal	Goal-acc Goal-acc Man Goal-gen	26,256 261 138,258
obnovit obnovení	renew, resume	Act Goal	Goal-acc Goal-gen	229 238

Figure 6 Czech: creative directed material processes

Wordbasis	Gloss	Theoretical valency frame	Realisation of Participants in the Corpus	Corpus References
editovat	edit	Act Goal	Goal-acc Man-instr	70,207,278
opakovat	repeat	Act Goal	Goal-acc	101
opsat opsání	write-around	Act Goal	Goal-dat	204
opustit opuštění	leave	Act Goal	Goal-gen	99
otevřít	open	Act Goal	Goal-acc Man-instr	79
pohybovat	move	Act Goal	Goal-instr Goal-instr Loc-na+acc Goal-instr Loc-podél+gen	226 268 270
pojmenovat	name	Act Goal	Goal-acc	131
pokračovat	continue	Act Goal	Goal-v+loc	235
posunout posunutí	shift, move	Act Goal	Goal-gen	86
použít	use	Act Goal	Goal-acc Goal-acc Goal-pro+acc	245 71
překřížit	cross	Act Goal	Goal-acc	112
přijímat	accept	Act Goal	Goal-acc	223
Rozkládat rozkládání	decompose	Act Goal	Goal-gen	73
spustit	start	Act Goal	Goal-acc Goal-acc Mans-instr	212 2,28,52, 195, 148
spustit	lower	Act Goal	Goal-acc Dir-adv	219

Wordbasis	Gloss	Theoretical valency frame	Realisation of Participants in the Corpus	Corpus References
stisknout stisknutí	enter, press	Act Goal	Goal-acc[1] Goal-gen	11,37,218 18
ukončit ukončení	end, terminate	Act Goal	 Goal-gen	 2,15,28,68
určit určení	specify, determine	Act Goal	Goal-acc Goal-gen	7,106,166,187 186,201
uzavřít uzavření	close	Act Goal	Goal-acc Goal-gen	142 175
vymazat vymazávat vymazání	delete, erase	Act Goal	Goal-acc Goal-acc Man-instr Goal-gen	271 279 263
vyplnit vyplnění	fill in	Act Goal	 Goal-gen	 128
vypsat vypsání	write-out, display	Act Goal	 Goal-gen	 156
začít	begin, start	Act Goal	Goal-acc[2] Loc-where Man-instr	17
začít	begin	Act Goal	Time	240
zadat	enter, specify	Act Goal	Goal-acc Goal-acc Goal-pro+acc[3] Goal-acc Condition-when Goal-acc Condition-if[4]	64,66 12,16 153 265
zarovnat zarovnání	align	Act Goal	 Goal-gen	 160
zaznamenat	record	Act Goal	Goal-acc	284
zmáčknout	press	Act Goal	Goal-acc	158,171

Figure 7 Czech: dispositive directed material processes

[1] The Patient of stisknout is either a name/label of a button or a key, or it is a specific description, e.g. tlačítko ENTER.

[2] The Patient of začít can also be realised by an infinitival complement, although no such occurrence has been encountered in the AGILE corpus.

[3] The Patient of *zadat* is either a specific value, *e.g. u, z, ?*, or it is a "generic" description, e.g. *vzdálenost (distance), úhel (angle), délka (length).* The purpose is never realised when the Patient is a generic description, When the Patient expresses a specific value the Purpose is realised in almost all cases.

[4] The condition is expressed by an if-clause.

Wordbasis	Gloss	Theoretical valency frame	Realisation of Participants in the Corpus	Corpus References
přidat přidání	add	Act Goal Dir	 Goal-gen Dir-“k”+dat Goal-gen	 85,90 137
uložit uložení	save	Act Goal Dir	Goal-acc Goal-gen Goal-gen Dir-“do”+gen	132,277 98 140
umístit	place	Act Goal Dir	Goal-acc Man-podél+gen	75
vepsat vepsání	write-into	Act Goal Dir	 Goal-elided Dir-“do”+gen	 189
vrátit vrácení	return	Act Goal Dir	 Goal-gen Dir-adv	 15,275
zapsat	write, save	Act Goal Dir	Goal-acc Dir-do+gen	231
zobrazit zobrazení	display	Act Goal Loc	 Goal-gen Goal-gen Loc-“v”+loc	 109,125

Figure 8 Czech: Directed material processes requiring a Spatial-locative complement.

Wordbasis	Gloss	Theoretical valency frame	Realisation of Participants in the Corpus	Corpus References
měnit, přeměnit, změnit přeměnění	change	Act Goal (Eff)	Goal-acc Goal-acc Time Goal-acc Eff-na+acc Goal-gen Eff-“na”+acc	72,74,282 72
přepínat přepnutí	switch	Act Goal (Orig) (Eff)	Goal-acc Man-pomocí+gen Eff-“do”+gen Orig-"z"+gen	247 59 249
spojit spojení	connect	Act Goal (Eff)	Goal-acc Goal-acc Eff Goal-gen,s+instr	114,118,121 113 14
vybrat výběr	select, choose	Act Goal (Orig)	Goal-acc Goal-acc Orig-z+gen Goal-acc Goal-pro+acc Goal-acc Goal-k+dat Goal-gen	84,89,107,124 4,81,152,214 84,89 137 96
vytvořit vytvoření	create	Act Goal (Ori)	Goal-acc Goal-gen	25,41 78,100
zvolit	choose	Act Goal (Orig)	Goal-acc Goal-acc Orig-z+gen	91,94,133 93,96,128

Figure 9 Czech: Directed material processes requiring a Spatial-locative complement

2.1.4.1.2 Observations in terms of valency frames

The mentioned observations can be conveniently recast in terms of valency frames. For an in-depth discussion of valency frames in the AGILE project context we refer to the LEXN2 deliverable of WP4; here we shall only recall some of the basic concepts.

Essentially, a valency frame for a word specifies how the word *considered as a head* can be modified, by listing the kinds of modifiers it can take. One dimension along which the modifiers of a head can be classified is whether they are **obligatory** or **optional**. Another dimension is whether a modifier is an **inner participant** (**IP**) or a **free modifier** (**FM**) (meaning whether the modifier can modify the head maximally once (IP) or possibly more than once (FM)). Contrastive to the former dimension, the distinction IP/FM is language-universal and thus irrespective of the head under consideration. The inner participants discerned in FGD are **Actor**, **Patient**, **Addressee**, **Effect**, and **Origin**. Characteristic to FGD's valency frames is that each valency frame contains *all* the inner participants, and specifies whether they are obligatory or optional, whereas free modifiers are (normally) only included in the valency frame if they are obligatory (Sgall et al. 1986).

In terms of SFG, Patient roughly corresponds to Actee, and Addressee to Beneficiary. As in FGD, they are both considered participants. However, a point of difference is the consideration of Effect and Origin: SFG does not consider *any* other participants besides Actor, Actee and Beneficiary. That is to say, from the viewpoint of transitivity anything else is regarded as a circumstantial.

Consequently, the Upper Model concepts defining the semantics of verbs in terms of processes do not contain semantic reflexes of Effect or Origin. Most importantly in the context of AGILE, **material processes** (reflecting an act of doing) are either **directed** or **nondirected**, depending on whether there is an Actee involved or not (to reflect *external agency* - see (Bateman et al, 1990)). Directed material actions are further divided into **dispositive actions** and **creative actions**, depending on whether the action *affects* the Actee or *brings about* the Actee.

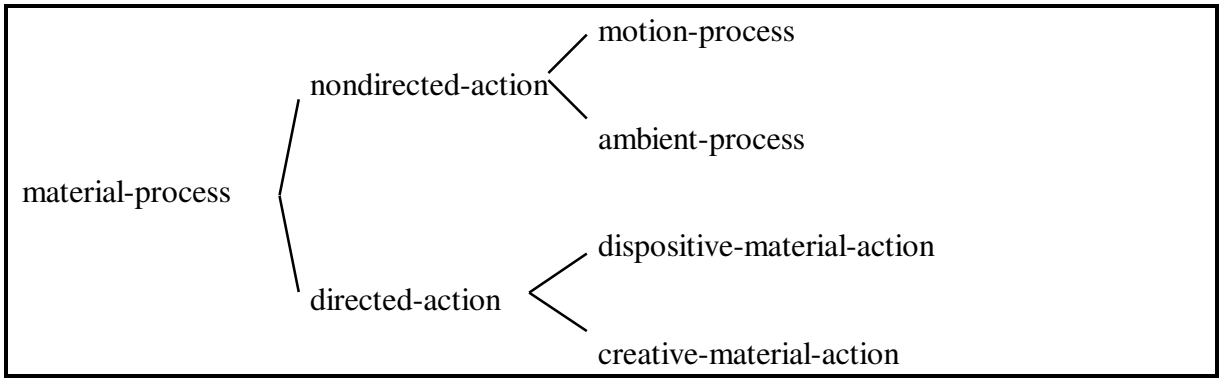

Figure 10 Material actions in the UM

However, as the data shows, some verbs include either an Effect or an Origin that can be argued for as an *inherent aspect of their underlying semantics*.

Translated into terms of the Upper Model, rather than viewing the (semantic reflexes of the) Effect or the Origin of a verb as circumstances *associated* to a process (modelling the semantics of the verb), we would argue that they be regarded as integral participants for specific classes of processes. For specific verbs occurring in the corpus they are therefore *not* to be considered as independent of the processes modelling the semantics of these verbs.

Thus, whenever an A-box would use the concept underlying one of these verbs, it should - by definition- also specify Effect, Origin, or both (depending on the concept's definition).

In the next subsection we consider this issue in some more detail.

2.1.4.2 Availability of Relevant Concepts in the UM

Before we continue our discussion, we would like to stress (again) that the issue we are raising here is relevant in the context of AGILE. The reason being that *all* the verbs in question (for example, *změnit*, *En.* to *change*) represent actions that are at the core of the AGILE domain.

To begin with, then, let us consider *Effect*, as in the following examples in which we use verbs, which occur in the Czech AGILE corpus:

(10) *Tento příkaz spojí čáry do párů*
This command conjoins lines into pairs

(11) *Tento příkaz vytváří obdélník z úseček*
This command creates a rectangle out of lines

(12) *Tento příkaz mění režim na Zarovnávání*
This command changes the mode to Justification.

In the UM, there is the **Cause-Effect** paradigm which models causal relations, in which there exists a cause, and an effect. The figure below shows Cause-Effect in more detail (cf. (Bateman et al, 1990), p.18ff). Of particular interest to us are the concepts (relations) **RST-volitional-result** and **RST-nonvolitional-result.** Both these relations concern volitionality of the action (or the absence thereof) - that is to say, whether the actor takes (or is to take) responsibility for the action (volitional) or not (nonvolitional). The first relation then expresses that the situation presented by the domain could be a cause for the volitional action, or a result of that action as presented in the range. The second relation indicates that the situation presented by the domain could be a cause for the nonvolitional action or result of a nonvolitional action as presented in the range.

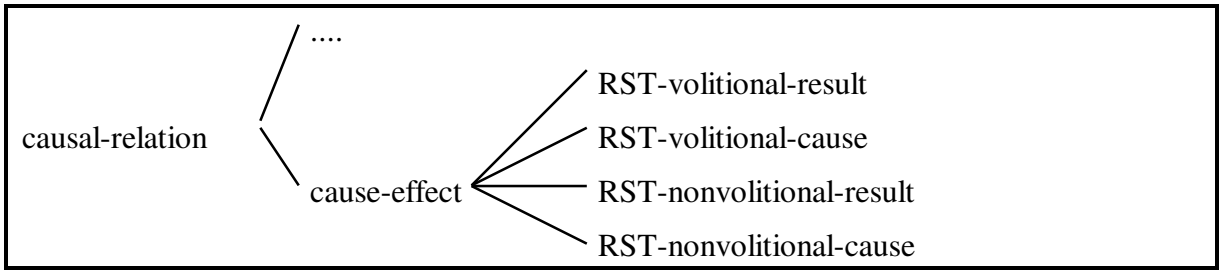

Figure 11 Cause-Effect in the UM

Within the context of AGILE, and particularly with regard to the observations stated earlier,

Effect can thus be perceived of as expressing an **RST-volitional-result**. In the next section we will present a proposal to extend the UM such that verbs, having as *partipant* an *Effect*, can be modelled by a process-concept that will require an **RST-volitional-result** to be expressed (generated).

Now, let us consider *Origin*. To begin with, *Origin* is not a spatial concept, (even though it may appear to suggest so at first sight), but should be conceived of in the light of the English

verb *to originate* - which means for example *to create, to come from, to spawn, to conceive*. We can exemplify this as follows. Consider the sentence in

(13) *Hiawatha made a canoe out of a tree.*

The constituent *out of a tree* is regarded as the *Origin* here, and it should be obvious that the tree provided the material out of which Hiawatha made the canoe. There is no spatial location involved[5]. Similarly in the following sentences containing verbs that appear in the Czech AGILE corpus:

(14) *Tento příkaz přepíná z režimu Krok do režimu Orto.*
This commmand switches from the Step mode to the Orto mode.

(15) *Tento příkaz vybírá z dostupných barev barvu elementu.*
This command selects element's color out of the available colors.

(16) *Nyní z nabízených možností zvolte typ zarovnání čar.*
Now select the type of line alignment out of the offered possibilities.

The seemingly close UM-concept **Source** is not an eligible candidate, since it is related to a process of motion, signifying where the motion is coming from.

The example in (13) also shows that *Origin* is neither a matter of extent: by the time that the canoe is finished, there is no longer something identifiable as a tree. The idea of transformation, change something (*Origin*) into something else (*Effect*), is almost inseparable from the idea of *Origin*. The UM-concept of **Material-Property-Ascription** thus comes to mind, which can have as property a **Material-Class-Quality**. This latter concept describes the quality of being made of a particular material, like wood or metal. For *Origin* we would need a similar concept, expressing that something is made *either* from a particular material, *or* arose (or was chosen) out of an object or set of objects.

2.1.4.3 Proposals

In order to bring about a *participant* Effect, we will explore in the implementation phase (WP 7) the possibility to relate **RST-volitional-result** to the meaning of a process – in particular, ensuring that the inquiry **:cause-condition-q** is called, and that it returns this value.

Bringing about a *participant* Origin requires a more thorough exploration. Under the UM-concept **Intension**, we would propose to replace **Material-Class-Quality** by two concepts, **Material-Intensional-Class-Quality** and **Material-Extensional-Class-Quality**, and ensure that the grammar (wherever relying on the original concept **Material-Class-Quality**) continues to work properly.

[5] We claim that it is rather unnatural to interpret the sentence as saying that Hiawatha made a canoe while coming out of a tree.

2.1.4.4 Types of Participants and Circumstances

Similarly to Russian, Czech is a highly inflectional language, in which grammatical relations are morphologically marked, as for instance reflected in a case government pattern. See the following example:

(17)

(a) *Příkaz OFFSET vytváří kopie čar.*
Command-nom *OFFSET creates*-3sg *copies*-acc *lines*-gen
The OFFSET command creates copies of lines.

(b) *Příkazem OFFSET se vytváří kopie čar.*
Command-instr *OFFSET* refl *create*-3sg *copies*-nom *lines*-gen
By the OFFSET command one creates copies of lines.

The process type is the same in both (a) and (b), and so is the structural configuration Actor—Process—Goal. In (a), which is in active voice, the Actor (*příkaz*) is conflated with Subject and carries nominative case and the Goal is conflated with the Direct Object carrying accusative case. In (b), the reflexive passive version of (a), the Goal is conflated with the Subject and thus carries nominative case and the Actor is expressed by a Circumstantial in Instrumental case.

In addition to the Actor, Goal and possibly Beneficiary, we suggest including Effect and Origin among the participants of certain verbs (see the discussion in the section on process types above).

Typically, the complementations of a verb in a clause are expressed by nominal groups that have the obligatory feature of CASE. One can distinguish, as in the KPML grammar, between the direct case functions (Subject realised by the nominative case, and Direct-complement realised by the accusative case) and the oblique case functions (others). The choice of the direct case functions is included in the voice system (see the next section on diathesis), choice of others depends on the process type and some rank conditions. In general the functions of Russian Case-Nominal-group are as follows:

Case	Basic meaning
Nom	when the Actor is conflated with the Subject (active voice), is it realised in the nominative case; always without a preposition
Gen	most often represents meta-actant relations, i.e. relations between nominalized process and its semantic actant, i.e. a Goal complement of a noun is mostly realised by a nominal group in genitive, e.g. the Goal in *kreslení čáry* (*the drawing of a line*); with or without a preposition
Dat	often the Addressee is realised by a nominal group in dative case; but also, some verbs take a Direct Object in dative in active voice, and some nouns take their Goal complements in dative; with or without a preposition, with preposition "k" means Direction to
Acc	many verbs realise their Goal in the accusative case when it is conflated with the Direct-Complement (active voice); with or without a preposition
Loc	combines with prepositions to express locations

Case	Basic meaning
Instr	expresses the Instrument; can realise the Agent, if the nominative-case is preselected for another nominal-group (complex passive); ; with or without a preposition

There are many meanings of cases and even more meanings of prepositions. Classes of verbs preselect complements realised in a particular way. We have discussed some of the possibilities of dealing with the valency frames of verbs in the lexicon and in the KPML grammar in the WP4 deliverable of Task 4.2.

2.1.4.4.1 Formal specifications of the nuclear transitivity

The initial model of TRANSITIVITY in the KPML grammar for Czech follows the proposal made for Russian (see above).

```
ACTOR-INSERT
{(MEDIATED, NONMETEOROLOGICAL, {INDICATIVE; JUSSIVE; IMPERATIVE-OTHER});
  (AGENT-INSERTED, MATERIAL)} →
(+ACTOR: NOMINAL-GROUP)

NONOBLIQUE-COMPLEMENTED
{MED-BEN; AG-BEN; NONBENEFACTIVE; NONCLIENCY; MENTAL; RELATIONAL; (MIDDLE
   VERBAL)} →
(ACTEE: ACCUSATIVE)

DEICTIC-SELECTION
{(MATERIAL-PROCESSUAL-AGENTIVE MATERIAL-PROCESSUAL);
  VERBAL-PROCESSUAL-AGENTIVE} →
[DEICTIC-AGENT] (AGENT/DEICTIC)
[NONDEICTIC-AGENT] (AGENT: NOMINAL-GROUP, INSTRUMENTAL)

NOMINAL-GROUP-CASE
nominal-like-groups→
   [nominative]
   [genitive]
   [dative]
   [accusative]
   [locative]
   [instrumental]
   [prepositional]
```

Figure 12 Initial model of TRANSITIVITY for Czech

Similarly to Russian, we have three basic types of phase processes: *opakovat* (repeat), *ukončit* (complete) and *pokračovat* (continue). In the current version, we follow the model proposed for Russian.

```
PHASE
TRANSITIVITY-UNIT →
[NOT-PHASE]
[PHASE] (+PHASE::PHASEVERB) (+PHASEDEPENDENT), (PHASE/VOICE),
    (PHASEDEPENDENT/LEXVERB)

PHASE-TYPE
PHASE → [BEGINNING] [REPEATING][ONGOING][ENDING]

BEGINNING-TYPE
BEGINNING → [STARTING] [BEGIN]

ENDING-TYPE
ENDING → [STOP] [FINISH]

ONGOING-TYPE
ONGOING →[KEEP] [CONTINUE]

PHASEDEPENDENT-TYPE
{BEGINNING; CONTINUE}→
[PHASEINFINITIVE] (PHASEDEPENDENT:::INFINITIVE)
[NOMINAL-GROUP-PHASE] (LEXVERB:NOMINAL-GROUP) (LEXVERB:ACCUSATIVE)
```

Figure 13 Initial model of phase type for Czech

We also need to alter the realisation of Thing in the nominal group rank, since when it expresses a nominalization of a process, its Goal is realised by a nominal group in a particular case, often but not always genitive case (see the tables in the corpus evidence section above).

2.1.4.4.2 Types of Prepositional Phrases in the corpus

In general, the complement of a verb can be expressed by nominal or adverbial groups. In this report we survey the prepositions and meanings of prepositional phrases as they appear in the Czech AGILE corpus. We do not provide here a full listing of all the prepositions available in Czech and the meanings of the prepositional phrases formed with them. Such classifications are available in the literature, e.g. (Šmilauer 1947). A listing of prepositions derived from adverbs or from noun is available within the context of the Prague Treebank Project (http://ufal.ms.mff.cuni.cz).

The AGILE corpus is specialised for a narrow technical domain and the repertoire of grammatical constructions found in it is rather limited. This concerns not only the list of prepositions occurring in the corpus as such, but also the range of the functions in which they are used. Mostly they are used to express Temporal, Spatial and Relational Circumstances. We encountered a frequent specific use of the preposition *pro (for)* in the analysed texts to express Purpose. The prepositional phrases with a nominalization correspond to the English *in order to* construction in our texts. This is a construction, which is becoming more and more common especially in technical texts, even though it is not really a proper Czech way of expressing Purpose. Czech would prefer a purpose clause.

- *pro (in order to, for)* + loc
  - in the analyzed text it is used to express purpose, the reasons for which the process is taking place; it is very often though not exclusively used with deverbal nouns (nominalizations):
    - *pro ukončení (for ending), pro uzavření (for closing), pro výběr (for selection)*

- *z (from)* + gen
  - as a circumstantial it is used in the corpus in Spatial Locatives expressing source with processes of taking something out of a collection or as a source with processes of taking something from the inside of something:
    - *jedním z příkazů, z plovoucího menu, z nástrojového panelu*
  - as a participant it serves to express Effect: ze segmentů
- *na (on)* + acc or + loc
  - as a Circumstantial with the accusative it expresses Direction somewhere, towards something:
    - *na konec čáry*
  - as a Circumstantial with the locative it expresses a Spatial Location on the surface of something:
    - *na nástrojovém panelu, na příkazové řádce*
  - as a Participant with accusative case it serves to express the Effect:
    - *zkonvertovat na segmenty, změnit na 0*
- *v (in)* + loc
  - expresses the Spatial Location of placement of something inside or at something:
    - *v dialogovém panelu. v obrazovém políčku, čáry ve vrcholech multičářry*
  - expresses a manner
    - *čáry v tečném směru*
- *do (to, into)* + gen
  - as a Circumstantial it expresses Direction into something or towards some goal
    - *návrat do dialogového panelu, uložení stulu do souboru, zapsánání čar do databáze*
  - as a Participant it can express Effect:
    - *conjoin elements to pairs*
- *pod (under)* + instr
  - expresses Spatial Location where, under which heading something is
    - *pod jménem, pod Popis*
  - direction or location lower than something
    - *pod originálem*
- *při (at)* + loc
  - condition or precondition; often with nominalization like *pro*
    - *při dotazu, při zakládání křivky. při pohybu zařízením*
- *od (from)* + gen
  - expresses direction away from the proximity of something
    - *od nové polohy kurzoru*
  - expresses manner
    - *kreslení od ruky*
- *po (on, after)* + loc
  - expresses Spatial Locative on the surface
    - *pohybovat kurzorem po obrazovce*
  - expresses Temporal circumstance, succession
    - *po provedení záznamu, po dokončení skicování*
- *s (with)* + instr
  - relational; expresses connection with something, appurtenance
    - *spojení prvního bodu s posledním, element s negativním offsetem*

- *k (for)* + dat
  - in our texts is expresses purpose
    - *vyberte Přidat k přidání elementů*
- *bez (without)* + gen
  - expresses exclusion, absence of something
    - *pohybovat kurzorem bez kreslení*
- *před (before)* + instr
  - inour texts is expresses a Temporal circumstance, precedence
- před kreslením nastavte proměnnou

2.1.5 Crosslinguistic summary description

The Slavonic languages under consideration are similar in the ways in which they structure the piece of reality reported in a clause in terms of process, participants and circumstances. These classes also closely follow the Upper Model. The ideational structure of the clause in modern Indo-European languages, in Slavonic languages, in particular, largely depends on the lexical semantics of the process (this is in contrast to ancient languages), cf. (Zimmerling, 1998). Clause structure also depends on typological characteristics of the language, including: obligatory *vs.* optional Subject; and presence *vs.* absence of empty pronouns filling vacant positions in the clause structure. For example, classes of meteorological processes or natural forces are often expressed by verbs without an actor: *It rains* (English); *вали* (Bulgarian); *Cz моросит* (Russian, lit. *It is drizzling*). Also in Slavonic languages a verbal root may be subjected to affixal modifications and alternations, influencing not only the lexical meaning of the resultant word, but also the ideational structure, for example, the participants possible for a predicate: *светает* (it is dawning), *светит солнце* (the sun shines), *солнце освещает дома* (the sun illuminates the houses). This also concerns aspectual pairs and nominalization.

The distinction between participants and circumstances in Slavonic languages follows the distinction drawn by the English grammar: participants are typically expressed by nominal groups in particular cases (though sometimes by prepositional phrases as well), while circumstances are typically expressed by prepositional phrases (or adverbial groups). In Czech and Russian the system of morphological case is used quite heavily for distinguishing types of processes and participants. In such general semantic functions as location or causation the languages are similar too, in the sense that they are realised by “universal” syntactic means such as prepositional phrases or conjunctions. However, these languages often have some additional means to convey the same oppositions as English: location, direction, orientation. The difference lies not in the inventory of syntactic functions, rather in their realisation in terms of rank.

The Russian and Czech languages have some peculiarities in comparison to English in the means of distinguishing between participants and circumstances. In English there are two criteria for participants: (1) it is arguable (can be a subject); (2) realised by NG - direct participation in a process (without preposition). In Russian and Czech according to the first criterion, the only subject roles are Actor and Goal, usually realised by the nominative case (Actor and Goal with passive form) and the accusative (Goal with active verb form). So, the amount of «participant part of a process» in Czech and Russian is narrower than in English. On the other hand, the two oblique cases — Dative and Instrumental — are used for marking participants, the former always coding Addressee and the latter Actor in the passive voice. Their function is similar to the function of role markers. There is no special case means in Russian to express Beneficiary. It is expressed by a PP with the preposition *для*. In

a prepositional phrase the semantic relation between a process and participants/circumstances is always expressed by a preposition - *для (for)* - Beneficiary, *в (in)* - Locative, *с помощью* (*by means of*) - Instrument and so on. However, the instrumental role can be expressed by a nominal group without a preposition: *выбрать мышью, выбрать с помощью мыши* (*select by the mouse*).

Cases preselected for the Minirange of a prepositional phrase either depend on the Minor-process (*с помощью* + genitive), thus they are completely formal. Cases may also reflect some inherent, grammaticalised characteristics of a process, for example, directness of the process – *в окно* (Accusative -> a movement: *towards a window*) - vs. *в окне* (Locative -> a stationary location: *in a window*). In this case the meaning of a prepositional phrase is independent from the lexical semantics of the predicate and is naturally treated as a "classic" circumstance. Since Bulgarian lacks morphological case, there are no significant differences in comparison to English, so the relation between a major process and a prepositional group is marked solely by the minor process.

2.2 Diathesis

Diathesis as a linguistic phenomenon provides the realisation of ideational process-participant configurations in clause, thus playing a distributing role rather than a classifying one. It has two facets, lexical and grammatical, which are distinguished in the process of wording as lexicalization (see transitivity above) and voice.

Voice is a grammatical facet of diathesis, so it serves to realise ideational structures presented as language oriented roles of Participants: Actor, Goal, Addressee and Beneficiary in the clause. Only material and verbal processes are fully relevant for voice transformations, mental and relational ones distribute their roles mostly by lexical means, converting roles of Senser and Phenomenon: cf. *like - please* (Halliday, 1985). On the other hand voice is strictly related to the grammar as a system of grammatical forms. Clause has some syntactically marked positions – grammatical functions involved in voice constructions, the three main ones being subject, direct complement and agentive complement. The following are examples of different realisations of the same ideational content taking into account active or passive modes for the delivery of meanings.

(1) *Bulgarian*

(c) *Командата* OFFSET *създава копия на линиите.*
Command-Def OFFSET creates *copies of lines*-Def.
OFFSET creates copies of lines.

(d) *С командата* OFFSET *се създават копия на линиите.*
With command-Def OFFSET *create*-refl *copies of lines*-Def.
By the OFFSET command copies of lines are created.

(2) *Czech*

(e) *Příkaz* OFFSET *vytváří kopie čar.*
Command-nom OFFSET creates *copies*-acc *lines*-gen.
OFFSET creates copies of lines.

(f) *Příkazem* OFFSET se vytváří *kopie čar..*
Command-instr OFFSET refl *create-3sg copies*-nom *lines*-gen.
By the OFFSET command copies of lines are created.

(3) *Russian*

(g) *Команда* OFFSET *создает копии линий.*
Command-nom OFFSET creates *copies*-acc *lines*-gen.
OFFSET creates copies of lines.

(h) *Командой* OFFSET *создаются копии линий.*
Command-instr OFFSET *create*-refl *copies*-nom *lines*-gen.
By the OFFSET command copies of lines are created.

The ideational structure in (a) and (b) can be treated in two ways. According to the first interpretation, it is constant in both (a) and (b): Actor—Directed-Material-Process—Goal. In (a), which is in the active voice, the Actor (*command*) is conflated with Subject and is expressed in the nominative case (in Russian and Czech) and the Goal is conflated with the Direct Complement expressed in the accusative case (in Russian and Czech). In (b), the passive version of (a), the Goal becomes Subject and thus is expressed by the nominative case (in Russian and Czech) and the Actor is expressed by the instrumental case.

According to the second interpretation, the command expresses the Instrument, which is conflated with the subject in (a) and is expressed as an instrument proper in (b), while the Actor (the doer) is implicit in this situation. The metaphoric power for reinterpretation of a command as a Doer rather than an Instrument for doing something is different in our languages. In English and Russian this reinterpretation is easy to achieve, so these languages readily transmit all the agentive characteristics to an inanimate object like a command, while Czech distinguishes between these roles, so **Čára se kreslí uživatelem* (*Lines are drawn by the user*) is impossible in Czech, while *Čára se kreslí příkazem Pline* (*Lines are drawn by the command Pline*) is OK, since the Actor in does not fit into the reflexive passive voice in Czech, while the Instrument does. The implications of these disparate interpretations will be addressed in more detail in the context of WP5.

Different diathesis perspectives on a process do not necessarily lead to a realisation by respective voice forms of a verb. One such example is the result of an ergative interpretation of processes in English clauses, when only the more inherent participant of a process is distinguished to realise the function of subject or direct complement (*the user starts the command - the command starts*) without any changes in the verb form. (For more details,

see the section on Contrastive analysis below.) Another example is the expression of multiple participants: *Connect the start point with the end point* vs. *Connect the start and the end points*. So diathesis is treated as the correspondence between the semantic roles of process participants and grammatical functions in a clause. Other constraints which influence the realisation of participants are contextual and lexical restrictions, such as the unavailability of reflexive or non-reflexive forms, or the impossibility of passive voice for intransitive verbs. Finally, there may be idiosyncratic conditions, such as the impossibility of the future form first person singular for the Russian verb *победить –*победю* (to win) while the first person plural form – *победим* and others exist.

2.2.1 Description for Russian

2.2.1.1 Theoretical background

Three voice forms are distinguished for verbs in the Russian grammar: active, passive and reflexive. They refer to different meanings depending on the context and lexical meaning of the verb. These forms as they are used in the Russian corpus for the intermediate demonstrator are the following:

1. active: *Команда OFFSET создает копии линий* [24] (*OFFSET creates copies of lines*)
2. passive: *После того, как многоугольник нарисован* [172], (lit. *After that, as polygon created*-perf; *After you've created a polygon*,...)
3. reflexive: *элементы отображаются в диалоговом окне* [83] (*Elements are displayed in the dialogue box*)

These forms refer to the three main voice meanings: active (1), passive (2), and middle (3) in accordance with the notions described above. In the passive voice the forms differ in relation to the aspect of the verb: the perfect verbs usually use the short form of passive participle (2), imperfect ones - reflexive forms with suffix *–ся*; respective imperfect passive voice forms (like *отображаем*) are not used in the modern language. The passive perfect form emphasises reached result, so it is not appropriate in discussion of possibilities (the passive in this example is incorporated in the subordinate clause), while the imperfect reflexive form refers to habituality of the action.

Traditional Russian grammars do not distinguish the middle voice as in English, in which the Medium is raised to the Subject position. A similar effect in Russian is achieved by reflexivisation, so that the Agent may be excluded from the participant set – *Пользователь запускает команду* (*The user starts the command*) vs. *Команда запускается* (*The command starts*). Since this meaning is close to the passive voice, it is labelled as *quasi-passive* in Russian. In quasi-passive there is a conflation of the semantic roles of Agent and Instrument, both of which are expressed in Russian by the instrumental case:

Командой OFFSET создаются копии линий
Command-inst *OFFSET create*-refl *copies of lines*
Copies of lines are created by the OFFSET command.

Another regular form of diathesis in Russian is nominalization with change of the direct complement of a verb to the genitive complement of a noun: *запускать команду* (*start command*-acc) vs. *запуск команды* (*starting command*-gen).

2.2.1.2 Evidence from the corpus

Active constructions.

The overwhelming majority of clauses in our texts are imperatives which lack an expression for the subject role and whose Actee is realised by a direct complement in the accusative case. All imperative clause in the text corpus are active; passive is not possible in the Russian imperative. All infinitive and gerund clauses are in the active voice too. Some explanations are given in the indicative mood in the active voice, in which the Actor represents a command and is conflated with the Subject role:

Команда OFFSET создает копии линий [24]

Passive constructions.

The passive voice in the perfective aspect is used in the subordinate clauses:

Введите с [168] для указания на то [169], что многоугольник описан вокруг круга.
Enter c [168] for reference to [169] that polygon circumscribed about circle.
Enter c for Circumscribed about circle.

После того, как шестиугольник нарисован [172], его можно редактировать с помощью команды PEDIT [173]
After that, as polygon created-perf, it-acc (Actee) can edit by means of command PEDIT.
After you've created a polygon, you can edit it with PEDIT.

Если перо было опущено, оно автоматически поднимется.
If pen was down, it automatically moves-up-refl.
If the pen was down, it moves up.

There are also passive participles functioning as heads of rank-shifted relative clauses:

последняя нарисованная линия [21] - the last line drawn
линии, смещенные на указанное расстояние [24] – lines offset at a specified distance
вновь созданный стиль мультилинии [102] - the newly created multiline style.

Quasi-passive constructions.

This voice type is preferred over the active voice in our corpus, since the Actor is not a conscious being. This allow it to be expressed in the instrumental case, which is a tool for realisation of both roles of Agent and Instrument:

Командой LINE создаются единичные или множественные линейные сегменты, [39]
Command-inst LINE create-refl single or multiple line segments-nom.
LINE creates single or multiple line segments.

Middle constructions

Constructions in reflexive form without Actor can be considered as Russian equivalents to the middle voice in English:

Этот цвет не показывается ... [92]
This colour-nom not show-refl ...
This colour does not show ...

Nominalizations

Nominal groups expressing processes are more frequent in the Russian corpus than in the English one; for example:

нажатием клавиши Return [19]
pressing (key-gen) Return

and so on.

2.2.1.3 Formal specifications

The first choice for English influencing the clause type, has a specific perspective in Russian: AGENCY is the distinction between middle (e.g. *the bomb exploded)* and effective (e.g. *the squad exploded the bomb).* The effective option represents the process as having a causer, the Agent, external to the participant present in the middle clause. As it follows from the section on transitivity, the majority of processes in this register are material-actions, which occur in active or in (quasi-)passive voice close to the English Middle semantically. So in our corpus a strict transitive model of the Russian verb does not affect the choice system. We can conserve the English model of grammar choices for voice, but change the realisation statements. Roles of Cliency and Recipiency are not used in our corpus. Choice of the subject depends on choices in the ideational (presence of Agent), textual (theme-rheme, topic-focus distribution), and interpersonal (assignment of modal responsibility) metafunctions.

The verb form in the passive voice is affected by the aspect: perfective verbs are mostly used in passive form - full or short passive participle, while the imperfective is in the reflexive. Generation of the Middle voice for Russian follow the following path in grammar:

```
PASSIVE-TYPE
{middle; passive}→
   [passive-habitual] (Process:::reflexive)
   [passive-perfective] (Process:::passive) (+Voice::bytj-aux)
                          Mood(Voice, Process)

AGENCY
{material; mental; verbal}→
   [middle]
   [effective] (Process::effective-verb)

MEDIUM-SUBJECT-CONFLATE
(medium-inserted, {(nonranged, middle); range-operative; medioreceptive;
    ({receptive; nonagentive}, {nonbenefactive; mental; relational})}) →
   (Medium/Subject) (Process:::reflexive)

PASSIVE-PROCESS
{range-receptive; receptive; medioreceptive; nonagentive; nonmediated}→
   (Voice::reflexive)
  [Passive]

EFFECTIVE-VOICE
(agentive, effective)→
   [operative] (Medium/Directcomplement)
   [receptive]

ADJUNCT-AGENT
{receptive; medioreceptive}→
   (Agent:::instrumental)
```

2.2.2 Description for Bulgarian

2.2.2.1 Theoretical background

Voice as a category has been the object of much heated dispute in the academic literature on the Bulgarian language. The major controversial issues include the number of voices in Bulgarian, i.e., two, three, or even five; the status of voice as a category, i.e., whether it is a morphological or a syntactic category; the nature of the «third» member of the category of voice, i.e., the so-called reflexive, or middle, voice. Some authors even question the adequacy of the theory on voice in Slavonic languages in general as it differs significantly from the original Old Greek and Latin paradigm on which the category of voice has been modelled (Stojanov, 1980).

By way of an example of the complexity of the problem, consider the following in which the reflexive verb *бия се* has two possible meanings—reciprocal action («fight») and action with a passive recipient («be beaten»):
(a) *Децата се бият на двора.*
children_Def fight_Refl in yard_Def
«The children are fighting in the yard.»

(b) *Децата не бива да се бият от родителите.*
children_Def not should to fight_Refl by parents_Def
«Children must not be beaten by parents.»

As these examples illustrate, the verb form itself remains the same (*се бият*). However, the meaning is different not only lexically but also, and more importantly, in terms of the description of the roles of the participants associated with the action referred to by the verb. Thus, in terms of the distinction between active and passive voice constructions, in example (a) the reflexive verb form can be interpreted as a kind of active construction (the children are Actors) whereas in example (b) the children are the recipients, Actee, or Goal of the action. Such usage of reflexive forms and similar considerations lead some authors to the conclusion that reflexive verb forms are not related to a specific voice but can be assigned to either active, or passive, depending on contextual, lexico-semantic and other factors.

2.2.2.2 Evidence from the corpus

Leaving aside much of the controversy around the status of voice in Bulgarian, it is possible to classify voice into three subcategories—active, passive, and reflexive. Reflexive clauses in their majority are a variant of passive constructions, which can also be considered as quasi-passive. These three types will be exemplified in the following from the Bulgarian corpus.

Active constructions
As in the Russian corpus, most clauses in the Bulgarian corpus are in the imperative mood, of positive polarity, and in the active voice used to convey instructions. There are a number of clauses in the indicative mood used for descriptive purposes, which are also predominantly in the active voice, for example:

Линиите пресичат края на цялата мултилиния, а външните дъги съединяват крайщата на най-външните елементи. (clauses 86-87 of the Bulgarian corpus)

Lines cross the end of the whole multiline and outer arcs join the ends of the outermost elements. (p. 48 of the English corpus)

Когато местите въвеждащото устройство, AutoCAD *чертае временни линейни сегменти с дължината, която вие сте му задали.* (clauses 181-182 of the Bulgarian corpus)

When you move your pointing device, AutoCAD draws temporary freehand line segments of the length you specified. (p. 52 of the English corpus)

Passive constructions

Passive constructions and verb forms are less frequent in Bulgarian, and in the corpus, respectively. Here are some of the occurrences in the corpus:

След като вече сте записали линиите, начертани на ръка, не можете да ги редактирате … (clauses 227-228 of the Bulgarian corpus)

Once you record freehand lines, you can't edit them … (p. 53 of the English corpus)

… ако всички въведени вече линии са записани … (clause 234 of the Bulgarian corpus)

… all lines entered so far have been recorded … (p. 53 of the English corpus)

Quasi-passive constructions

Finally, there is a number of constructions in the corpus which can be called quasi-passive and have similar functions as in the Russian text with the exception of the different usage of cases in Russian and their absence in Bulgarian. The following examples illustrate their usage:

Елементите се появяват в диалоговия прозорец (clause 79 of the Bulgarian corpus)

Elements are displayed in the dialog box (p. 47 of the English corpus)

… ако искате да се появи линия във върховете на мултилинията … (clause 83 of the Bulgarian corpus)

… to display a line at the vertices of the multiline … (p. 48 of the English corpus)

2.2.3 Description for Czech

In this section, we first describe the system of voice in Czech in general, and its uses in diathesis, in particular for the sake of «deagentization» of clauses where the Actor does not function as the Subject. We shall concentrate on the cases where Subject is either left out or is conflated with Goal, which thus becomes foregrounded. Then we provide evidence of voice usage from the Czech AGILE corpus. In the last subsection we propose a formalisation of voice for the Czech KPML grammar. In this deliverable, we do not discuss

other types of deagentization or diathesis in more general, as they are discussed in some literature, e.g. (Daneš et al. 1987, Karlík et al.1995).

2.2.3.1 Theoretical background

Our description of voice in Czech is based on the Czech grammar description in (Šmilauer 1947; Panevová, p.c.). Voice (*Cz.* slovesný rod, genus verbi) is considered to reflect the relationship between the experiential functions of the Actor and the Goal of a process and the grammatical structure of the clause, namely the choice of the Subject and Object (Direct Complement). By Actor (agents), one means the agent from whom the process originates, who brings the process about. Whereas the Goal (patients) is the agent at whom the process is directed or who is affected by it. In the example below, 'user' is the Actor and 'line' is the Goal.

(1)

(a) *Uživatel* *kreslí* *čáru.*
User-nom(Subject/Actor) *is-drawing* *line*-acc(Object/Goal).
The user is drawing a line.

(b) *Čára* *je kreslena* *uživatelem*
Line-nom(Subject/Goal) *is being-drawn* *user*-instr
A line is being drawn by the user.

(c) *Čára* *se* *kreslí.*
Line-nom(Subject/Goal) refl *is-being-drawn*
A line is being drawn.

Two main types of voice are distinguished in Czech: **active** and **passive**. In a clause in active voice, the Actor fulfills the function of Subject, see (1a) above. In a clause in the passive voice, the function of Subject is not fulfilled by the Actor. There are two forms realizing the passive voice: **complex passive** and **reflexive passive**, see (1b) and (1c) above, respectively. The complex passive is formed by the auxiliary verb *být (be)* and a passive participle. Furthermore, **personal passive** and **impersonal passive** is distinguished. A clause in personal passive includes a Subject, as in (2), whereas a clause in impersonal passive does not, as in (3).

It should be noted that a clause does not always contain an explicit Subject. Due to the pro-drop character of Czech, the Subject can be dropped when the following conditions hold:

- the verb ending sufficiently determines the Person (first person: the speaker, second person: the hearer), or it is recoverable from the context (third person)
- Person is sufficiently determined by the verb ending, the meaning of the Subject is clear and dropping the Subject does not counter the communication purpose

- the verb ending sufficiently determines the Person (first person: the speaker, second person: the hearer), or it is recoverable from the context (third person)
- Person is sufficiently determined by the verb ending, the meaning of the Subject is clear and dropping the Subject does not counter the communication purpose

In other words, Subject can be dropped in Czech if the conditions hold for its realisation by a weak pronoun, c.f. examples (2a) and (2b) below. In the examples below we mark a dropped Subject by «∅» when needed. For more discussion of the possibilities of realising the Actor in Czech see Section 2.2.3.1.3.

Both active and passive voice are generally possible in indicative, conditional as well as imperative mood. Passive imperative is uncommon and very literal, not suitable for technical texts. The usage of active and passive voice in indicative and conditional mood follow the same pattern. All our examples in this theoretical section will be in indicative.

(2)

(a) *Čára je kreslena uživatelem*
Line-nom(Subject/Goal) is being-drawn user-instr
A line is being drawn by the user.

(b) *Je kreslena uživatelem*
Is being-drawn user-instr
It is being drawn by the user.

(c) *Čára se kreslí.*
Line-nom(Subject/Goal) refl *draw*-3sg
A line is being drawn.

(d) *Kreslí se.*
draw-3sg refl (dropped Subject/Goal)
It is being drawn.

(e) *Vždy bude vzpomínán.*
Always will-be-3sg *remembered*-3sg-masc (dropped Subject/Goal)
He will be always remembered.

(f) *Kreslí se.*
Draw-3sing refl
Drawing is going on.

(g) *Vždy ho bude vzpomínáno.* (Šmilauer 1947)
Always he-gen *will-be*-3sg *remembered*-3sg-neut
One *will always remember him.*

(Šmilauer 1947) distinguishes several factors that condition the usage of active versus passive voice. We adopt his classification here, but we replace some of the occurrences of the terms Subject and Object in his classification by Actor and Goal, respectively. The factors important for the use of active versus passive voice are thus the following:

- whether the verb is
  - impersonal (without Actor, with or without Goal)
  - personal
    - with Actor, without Goal, so-called Subject verbs in (Šmilauer 1947), we shall call them **verbs with no Object**
    - with Goal, which when it is conflated with Object is expressed
      - in accusative case, so-called Transitive verbs in (Šmilauer 1947), we shall call them **verbs with Object in accusative**

  - by other means, so-called Non-transitive verbs in (Šmilauer 1947) (note: these are not intransitive verbs!), we shall call them **verbs with non-accusative Object**
- whether the Actor is
  - explicitly realised
  - implicit (not realised)
- what is the relationship between the Subject and the Actor and Goal

Let us now discuss the usage of voice in Czech depending on these factors. Our discussion is based on (Šmilauer 1947).

2.2.3.1.1 Impersonal verbs

Impersonal verbs express processes without an Actor. The clauses with impersonal verbs do not contain a Subject. An impersonal verb can express either a process with a Goal or a process without a Goal.

a) when the verb expresses a process with a Goal, it combines with an Object, e.g. *zebe mě* (*I am cold*)
b) when the verb expresses a process without a Goal, it does not combine with an Object, e.g. *sněží* (*it is snowing*).

Impersonal verbs do not differentiate for voice. Their verb form is by definition either active, e.g. *zebe mě* (*I am cold*), *sněží* (*it is snowing*), or (lexically) reflexive, e.g. *chumelí se* (*it is snowing*), *stýská se mi* (*I am lonesome*).

2.2.3.1.2 Personal verbs

Personal verbs express processes with an Actor. The Actor can be realised explicitly in a clause, or it can be implicit (not realised). With personal verbs, the issue of voice choice arises, namely whether the Actor is or is not realised by the Subject of the clause. As we noted above, due to the pro-drop character of Czech, if a clause has a Subject, it can be explicit or dropped (marked by «∅» in the examples, for clarity).

In the following discussion, we distinguish among personal verbs according to whether they express processes without a Goal or with a Goal. In the latter case, we further distinguish whether the Goal can be realised as an Object in accusative or not.

2.2.3.1.2.1 Personal Verbs with no Object

Verbs without an Object express processes without a Goal, namely states or state transitions: *ležet (lie), sedět (sit), spát (sleep), padat (fall), růst (grow), chybět (miss)*, or the lexically reflexive ones *smát se, bát se*; the reflexive verb *objevit se (appear)* which occurs in the AGILE corpus is of this type, too[6]. In a clause, these verbs can combine only with a Subject, not with an Object. They can be used in the following ways:

[6] Note that the nonreflexive verb *objevit* (discover) is transitive.

a) with Subject: the Subject realises the Actor of the process, the clause is in active voice, e.g. *Dědeček|∅ chodí denně na procházku (Grandpa|He goes daily for a walk) Pero|∅ klesne dolů (The pen|It lowers down) Objeví se okno|∅ (A window|It appears)*
b) without Subject (impersonal use, Actor is unspecific): the clause is in passive voice, namely in reflexive passive, e.g. *Chodí se denně na procházku (One goes daily for a walk) Kleslo se dolů (One lowered down)*; the complex passive is possible, but very rare in contemporary Czech. The passive cannot be formed for the reflexive verbs, like *objevit se (appear)*, e.g. **Objevilo se*[7] **Bylo se objeveno*.

2.2.3.1.2.2 Verbs with non-accusative Object

Verbs with non-accusative Object can express processes with a Goal. In clauses they can be combined with a Subject and with an Object which is expressed by other means than a nominal group in the accusative case. They can be used in the following ways:

a) with Subject: the Subject realises the Actor of the process, the clause is in active voice
   (i) with Object: the Object realises the Goal, e.g. *Petr|∅ odpověděl na otázky (Petr|He replied to questions)*
   (ii) without an Object: e.g. *Petr|∅ většinou odpověděl (Petr|He usually replied)*
b) without Subject: the clause is in passive voice
   (i) with Object: the Object realises the Goal; either the reflexive passive or the complex passive can be used
      * reflexive passive, e.g. *Odpovídalo se na otázky (One was repying to questions)*
      * complex passive, e.g. *Bylo odpovídáno na otázky (One was replying to questions)*
   (ii) without Object: reflexive passive is used, e.g. *Odpovídalo se (One was replying)*

For verbs with non-accusative Object, the complex passive is almost exclusively the impersonal type. Some rare exceptions can be found, for instance *Tím byl světový obchod silné otřesen* (*By this, the world trade was severely shaken*).

2.2.3.1.2.3 Verbs with Object in accusative

Verbs with Object in accusative also express processes with a Goal. In clauses, they can be combined with a Subject and with an Object, which is expressed by a nominal group in the accusative case. Their usage depends on whether the clause expresses both Actor and Goal, only Goal or neither Actor nor Goal. We discuss these uses below:

a) both Actor and Goal are realised
   (i) the Subject realises the Actor of the process: the clause is in active voice, e.g. *Uživatel|∅ kreslí čáru (The user|He is drawing a line)*
   (ii) the Subject realises the Goal: the clause is in personal passive
      * complex personal passive, e.g. *Čára|∅ je kreslena uživatelem (A line|It is being drawn by the user)*

[7] *Objevilo se* is a grammatical sentence in Czech. It can only be interpreted as Subject/Actor, with implicit Subject (dropped anaphoric pronoun).

b) only the Goal is realised, and it is conflated with the Subject: the clause is in personal passive voice, either reflexive or complex
   (i) personal reflexive passive, e.g. *Čára|∅ se kreslí (A line|It is being drawn)*
   (ii) personal complex passive, e.g. *Čára|∅ je kreslena (A line|It is being drawn)*
c) neither Actor nor Goal are realised: the clause is in impersonal reflexive voice, e.g. *Kreslí se (Drawing is going on)*

In these examples, we only included Actor or Goal, or none of them. It is of course possible to included another element, for instance an Instrument. This is a configuration often encountered in the AGILE register. Then we get the following possibilities:

a) Actor, Goal and Instrument are realised
   (i) the Subject realises the Actor of the process: the clause is in active voice, e.g. *Uživatel|∅ kreslí čáru příkazem* PLINE *(The user|He is drawing a line by the* PLINE *command)*
   (ii) the Subject realises the Goal: the clause is in personal passive
      * complex personal passive, e.g. *Čára|∅ je kreslena uživatelem příkazem* PLINE *(A line|It is being drawn by the user by the* PLINE *command)*
b) only the Goal is realised, and it is conflated with the Subject: the clause is in personal passive voice, either reflexive or complex
   (i) personal reflexive passive, e.g. *Čára|∅ se kreslí příkazem* PLINE *(A line|It is being drawn by the* PLINE *command)*
   (ii) personal complex passive, e.g. *Čára|∅ je kreslena příkazem* PLINE *(A line|It is being drawn by the* PLINE *command)*
c) neither Actor nor Goal are realised: the clause is in impersonal reflexive voice, e.g. *Kreslí se příkazem* PLINE *(Drawing by the command* PLINE *is going on)*

We propose to treat the instrument by which the process is done as the Instrument Circumstance in all the cases, no matter whether the Actor is realised or not. The reflexive passive construction in Czech does not allow the inclusion of an Actor: the sentence **Čára se kreslí uživatelem (A line is drawn by the user) is ungrammatical.*

We also propose to treat 'command' as an Instrument in the sentences such as (3a). (3b) is excluded because when we have a nominal group in nominative case and a nominal group in accusative case it is not possible to interpret the former as the Instrument and the latter as the Actor.

(3)

(a) *Příkaz* PLINE *kreslí čáru.*
Command-nom(Subject/Instrument) PLINE *is-drawing line*-acc(Object/Goal).
The PLINE *command is drawing a line.*

(b) **Příkaz* PLINE *kreslí čáru uživatelem.*
Command-nom PLINE *is-drawing line*-acc user-instr.
The PLINE *command is drawing a line by a user.*

The clauses where both Actor and Goal are realised provide the clearest opposition between the active and passive voice: in the active voice, the Actor is conflated with the Subject and the Goal with the Object, as in (4a), while in the passive voice, it is the Goal that is conflated with the Subject and the Actor can be expressed as a Circumstantial, as in (4b).

(4)

(a) *Uživatel* *kreslí* *čáru.*
User-nom(Subject/Actor) *is-drawing* *line*-acc(Object/Goal).
The user is drawing a line.

(b) *Čára* *je kreslena* *uživatelem*
Line-nom(Subject/Goal) *is being-drawn* *user*-instr
A line is being drawn by the user.

In languages like English the choice between active and passive voice can be motivated by the need of information structure, namely to achieve the order with the realisation of Goal before the realisation of the Actor, so that the Actor is more communicatively dynamic than the Goal. In Czech, however, as well as in Russian and other Slavonic languages, such ordering can be achieved simply by a change in word order, without a change in grammatical structure, as in (5).

(5) *Čáru* *kreslí* *Uživatel.*
Line-acc(Object/Goal) *is-drawing* *user*-nom(Subject/Actor).
A line is being drawn by a user.

The personal complex passive of verbs with accusative Object with both Actor and Goal realised is generally disliked in Czech. When it is used, the motivation often lies in textual cohesion, e.g., when a number of successive processes have the same Goal. When this Goal is conflated with the Subject, the predicates can constitute a complex predicate, e.g. *Kružnice byla zadána dvěma body, umístěna na ploše příkazem MOVE, vyplněna barvou pomocí příkazu FILL a příkazem SAVE uložena (The circle was specified by two points, placed in the area by the MOVE command, filled using the FILL command and saved by the SAVE command)*. In complex passive, the same Subject can be maintained. In the active voice, there would have to be multiple clauses with different Subjects.

As mentioned above, the personal complex passive of verbs with accusative Object with both Actor and Goal is acceptable when the Actor can also be understood as an Instrument.

Not all verbs with Object in accusative are able to form the passive form. Examples are *tato úsečka přesahuje okraj papíru (this line exceeds the paper margin)*, *to znamená konec (this means the end)*. Of all verbs occurring in the AGILE corpus, passive forms can be created.

Whereas the passive voice is quite rare in Czech in realizing processes where both the Actor and Goal are realised as mentioned above, it is rather common when only the Goal is realised. This is a regular type of the passive, especially the reflexive passive. The Actor is not realised either because it is unknown, unspecific, general or it is unimportant in the

speech situation. For example, *Zadávání objektů se ukončuje stisknutím ENTER (Specifying objects is ended by pressing ENTER).* Sometimes the Actor is left out for reasons of modesty.

When a verb with Object is to be used without realising the Actor, but with realising the Goal, there is a competition between the reflexive and the complex passive, e.g. *Čára je kreslena* or *Čára se kreslí (A line is being drawn), Na otázky se odpovídalo* or N*a otázky bylo odpovídáno* (*One was replying to questions*). There is not clear discrimination between these two ways in Czech, but (Šmilauer 1947) mentions the following tendencies:

a) Since the reflexive passive appears closer to the active, complex passive is preferred in contexts where the notion of passive undergoing (being subjected to some process rather than carrying out a process) is intended. Conversely, the reflexive passive is used when the notion of undergoing is not (or less) applicable.
b) Complex passive is rare in colloquial or informal language.
c) Reflexive passive is not used at all for Goals corresponding to or including the Speaker or the Hearer, because there would arise a collision with the other meanings of reflexives, namely reciprocal processes or process directed at oneself; e.g., *Nakreslila jsem se* can only mean *I drew myself* and not *I was drawn*, while *Čára se nakreslila* can mean *A line was drawn* as well as *A line drew itself*, but the latter is unlikely.
d) Since reflexive forms of verbs have also the passive function, one does not use complex passives created from active reflexives.
e) Complex passive forms are static, close to constructions expressing a state. In present tense the is a clear difference between complex passive and an adjective expressing state, e.g. *Soubor je ukládán* (*A file is being saved*, complex passive, activity carried out now) vs. *Soubor je uložen* (*A file is saved*, copula with a state adjective expressing a state or a property). However, in past tense, the expression *byl uložen* (*was saved*) can be analyzed either as a complex passive or as a copula with a state adjective. There is no clear boundary between the two cases. Therefore, the complex passive is used especially in contexts where there is a notion of a state resulting from some process, in particular with perfective verbs (u dokonavých sloves) and certainly in past tense.
f) Reflexive passive, on the other hand, is more dynamic than the complex passive. It is used especially for repeated, regular, habitual or typical activities. Compare: *Čára je nakreslena z bodu A do bodu B* (*A line is drawn from point A to point B*) vs. *Čára se kreslí z jednoho bodu do jiného bodu* (*On draws line from one point to another point*)

2.2.3.1.3 Realisations of Actor in Czech

Actor can be realised or not realised in Czech. The ways of realising an Actor are as follows:

- Actor/Subject: active voice
  - explicit Subject, e.g. *Příkaz LINE kreslí čáry (The LINE command draws lines)*
  - implicit (dropped) Subject, e.g. *Kreslí čáry (It draws lines)*

- Actor/Circumstantial in instrumental case: passive voice
  - complex personal passive voice, e.g. *Čáry|∅ jsou kresleny příkazem LINE (Lines are drawn by the LINE command), Čáry|∅ jsou kresleny uživatelem (Lines are drawn by the user)*
  - reflexive personal passive voice, only for Instrument-like Actors, e.g. Čáry se kreslí příkazem LINE (Lines are drawn by the LINE command), Jsou kresleny příkzem LINE (They are drawn by the LINE command)

Actor can be not realised when it is general, unspecific. It is possible that the speaker has some animate, underspecified Actor in mind, for instance a person in general, or some group of people. In the AGILE domain, this can be the user(s) of the system. Clauses with **general Actor** are either in active voice or in passive voice. The possibilities in indicative mood are as follows:

- without Subject
  - active voice
    - speaker included: Predicate in first person plural, e.g. *Čáry kreslíme příkazem LINE (We drawn lines by the LINE command), V novinách píšeme … (In the newspaper we write …)*
    - *speaker excluded, addressee included*: Predicate in second person
      - singular addressee, informal form, e.g. *Čáry kreslíš příkazem LINE (You draw lines by the LINE command), ?V novinách píšeš … (In the newspaper you write …)*
      - singular adressee, informal form or plural addressee, e.g. *Čáry kreslíte příkazem LINE (You draw lines by the LINE command), V novinách píšete … (In the newspaper you write …)*
    - *both speaker and hearer excluded:* Predicate in third person
      - third person singular, mostly idiomatic expressions of comparison and measure
      - third person plural, e.g. *V novinách píšou … (In the newspaper they writte*
  - impersonal passive voice
    - impersonal reflexive passive, e.g. *Čára se zruší, <u>když se klepne na CANCEL</u> (Line is cancelled when one klicks on CANCEL), <u>Když se stiskne ENTER</u> obrázek se uloží (When ENTER is pressed, the picture is saved), K velikosti objektu se nepřihlíží (The size of the object is not taken into account), V novinách se píše … (In the newspaper one writes …)*
    - impersonal complex passive, which is less common
      - with personal verbs without Object very rare
      - with personal verbs with non-accusative Object more common, e.g. *Na dotaz bylo odpovězeno (An inquiry has been replied to), Bylo použito nejnovějších metod (The newest methods have been made use of), V novinách je psáno … (In the newspaper is written …)*

- with Subject
  - active voice, e.g. *Člověk čáry kreslí příkazem LINE (Man draws lines by the LINE command), Noviny píšou … (The newspaper writes …)*
  - personal passive voice
    - personal reflexive passive, e.g. *Čáry se kreslí příkazem LINE (Lines are drawn by the LINE command), Když se stiskne ENTER obrázek se uloží (When ENTER is pressed, the picture is saved), Velikosti objektu se neuvažuje (The size of the object is not considered)*
    - personal complex passive with personal verbs with accusative Object, e.g. Dotaz byl zodpovězen (An inquiry has been answered), Byly použity nejnovější metody (The newest methods have been used)

Note that while the complex passive can be combined with an Actor (realised as a Circumstantial in instrumental case), the reflexive passive cannot be combined with an Actor at all.

2.2.3.2 Evidence from the corpus

2.2.3.2.1 Active constructions.

Majority of clauses in the Czech AGILE corpus are in the active voice. Many of them are imperatives for which passive form is not suitable in technical texts. In the imperatives, Actor/Subject is implicit and the Actee is realised by an Object (direct complement), mostly in the accusative case. Most other clauses are in the active voice, too. In some explanations in indicative mood in the active voice, the Actor conflated with the Subject is Instrument-like and realises a command, as exemplified (6).

(6) *TYPČ nastavuje aktuální typ čáry.* [26]
Command LINETYPE sets the current line type.

There are the two following occurrences of the reflexive impersonal verb *objevit se* (appear), which cannot be passivized, in the corpus:

(7) *Například, jestliže vytvoříte element s negativním offsetem, objeví se pod originálem v obrazovém políčku.* *[103-104]*
For instance, if you create an element with a negative offset, it appears under the original in a window.

(8) *Tato barva se neobjeví v obrazovém políčku v dialogovém panelu* Styly multičár. *[126]*
This colour does not appear in a window in the dialog box Multiline Styles.

2.2.3.2.2 Passive constructions.

Personal verbs with accusative Object, impersonal reflexive passive type

(9) *Při rozkládání široké křivky se šířka čáry změní na 0* *[73-74]*
When a polyline is decomposed the line width changes to 0.

(10) *Stiskněte znova výběrové tlačítko a pero se zdvihne tak, že nyní můžete pohybovat kurzorem po obrazovce bez kreslení.* *[224-227]*
Press again the choice button and the pen lifts so that you can now move the cursor across the screen without drawing.

(11) *Jestliže bylo pero dole, zdvihne se nahoru.* *[266-267]*
If the pen was down, it lifts up.

These are processes where it is difficult to decide whether they really are cases of reflexive passive constructions or whether the verbs are in fact "secondary" (derived) reflexive verbs. Even though there appears to be an original transitive verb from which the reflexive form could have been obtained by passivization, these all processes which can be perceived of as "self-driven". In the latter perspective, there would be no Actor, and thus no passive. This type is at the borderline between active and passive. In the linguistic literature on Czech voice it is referred to as the mediopassive.

The two views, i.e. passive vs. mediopassive, are competing. Both are possible and deciding between them is a matter of adopting a particular theory. Currently we decided to treat the above sentences as passive. This view does not preclude a later switch to mediopassive, if we include such possibility in the grammar and classify the verbs in the lexicon as for whether or not they can form the mediopassive construction.

In (9), the active form is *někdo změní něco na něco (someone changes something to something)*, he passive reflexive form is *změnit se na něco (change-refl to something).* In (10) and (11), the active form is *někdo zvedne něco (someone lifts something)*, the passive is *něco se zvedne (something lifts-refl).*

Personal verbs with accusative Object, impersonal complex passive type

(12) *Element nemusí být nakreslen v počátku.* *[88]*
The element does not have to be drawn in the beginning.

(13) *Zadejte kdykoli* ***z*** *a čára, kterou právě kreslíte, a ty, které již byly nakresleny, budou zapsány do databáze.* *[230-233]*
Enter z any time a the line you are just drawing and those already drawn will be entered into the database.

(14) *Elementy budou zobrazeny v dialogovém panelu Styly multičáry.* *[102]*
The elements will be displayed in the Multiline Styles *dialog box.*

(15) *Stiskněte opět tlačítko a kreslení bude obnoveno od nové polohy kurzoru.* *[228-229]*
Press again the button and drawing will be resumed from the new cursor position.

(16) *Aktuální pohled můžete měnit během skicování, jen pokud je pero nahoře, všechny čáry byly již zaznamenány a režim tabletu je vypnut.* *[282-285]*
You can change the current view during sketching only if the pen is up, all lines have been saved and the tablet mode is switched off.

All tenses are used, only perfective verbs are used. The active forms are *někdo nakreslí něco (someone draws something)* (12, 13)*, někdo zapíše něco (someone enters something)* (13)*, někdo zobrazí něco (someone displays something)* (14)*, někdo obnoví něco (someone resumes something)* (15)*, někdo zaznamená něco (someone records something)* (16)*, někdo vypne něco (someone turns-off something)* (16).

There are also passive participles functioning as heads of rank-shifted relative clauses:

(17) *úhel vytvořený pomocí … (the angle created using …)* [48]

(18) *k přidání vytvořeného stylu multičáry (for adding the created style of multiline)* [137]

2.2.3.3 Formal specifications

The first choice for English influencing the clause type is AGENCY, as the distinction between middle (e.g. *the bomb exploded)* and effective (e.g. *the squad exploded the bomb).* The opposition between effective and middle in the Czech active voice does not exist. But we can, in the initial stage, try and match the effective voice with the active and the complex passive, and match the middle voice with the reflexive passive. The effective option represents the process as having a causer, the Agent, external to the participant present in the middle clause.

Following the proposal made for Russian, we conserve the English model of grammar choices for voice, but change the realisation statements. Roles of Cliency and Recipiency are not used in our corpus. Choice of the subject depends on choices in the ideational (presence of Agent), textual (theme-rheme, topic-focus articulation), interpersonal (assignment of modal responsibility) metafunctions.

The initial formalization of Czech voice is shown in

```
PASSIVE-TYPE
{middle; passive}→
   [passive-habitual] (Process:::reflexive)
   [passive-perfective] (Process:::passive) (+Voice::byt-aux)
                         Mood(Voice, Process)

AGENCY
{material; mental; verbal}→
   [middle]
   [effective] (Process::effective-verb)


MEDIUM-SUBJECT-CONFLATE
(medium-inserted, {(nonranged, middle); range-operative; medioreceptive;
    ({receptive; nonagentive}, {nonbenefactive; mental; relational})}) →
   (Medium/Subject) (Process:::reflexive)

PASSIVE-PROCESS
{range-receptive; receptive; medioreceptive; nonagentive; nonmediated}→
   (Voice::reflexive)
  [Passive]

EFFECTIVE-VOICE
(agentive, effective)→
   [operative] (Medium/Directcomplement)
   [receptive]


ADJUNCT-AGENT
{receptive; medioreceptive}→
   (Agent:::instrumental)
```

Figure 14 Initial formalization of the Czech voice system

2.2.4 Crosslinguistic summary description

Traditional typological linguistics distinguishes two language systems – languages of the nominative type and languages of the ergative type. Languages of the nominative type maintain an Actor / Goal opposition and split all verbs into two lexical classes of transitive / intransitive verbs. Also nominative type languages provide an opposition of direct / oblique cases; among the former, the nominative case is used for marking subjects, and the accusative case for marking direct complements. Basically active / passive transformations are possible only for transitive verbs, in which the Goal is treated as a participant undergoing a manipulation by the Actor, however, sometimes (in particular, in Czech in our project), some verbs which have the ideational Goal role still are not considered as transitive ones, since the Goal is not expressed by the accusative case (for example, *odpovídat na otázky*, *answer to questions*). Ergative type languages maintain an opposition between Agent and Factitive (action carrier, or, in Halliday terms, the more inherent participant of the process); Agent is marked in some way (usually by the ergative case as, for example, in some Caucasian languages) and opposed to the Absolute case, unifying roles of Actor of an intransitive verb and Goal of a transitive verb. Halliday usually opposes transitive / ergative models of process, not nominative / ergative language types, though English is of nominative type but with the clear movement to ergative one. We shall use Halliday terminology: Actor - Goal and Agent - Medium to refer to transitive and ergative functions respectively.

Among languages investigated in our project, Russian and Czech have the obligatory case specification of a nominal group and belong to the nominative language type, so roles of participants are realised by cases, Bulgarian and English lack case markers for nominal groups and distinguish participant roles by their positions. If there are two of the more inherent participants (Actee and Actor), or only one participant (Actor), the same transitive/non-transitive pattern is followed. However, if only Actee is included in the participants set without an Agent, this leads to reclassification of syntactic functions in realisation. In this case the situation is similar to that in languages of the ergative type in which the nominative case is not a reserved position for Subject-Actor, however, Bulgarian and English lack a special ergative marker for the Agent. Status of Actor and Actee in this situation is conflated in one function of the more inherent participant in the process – Medium in SFG terminology. In this case (for nominative type of language) Halliday uses the term "Middle-voice", for example, *The command starts* (instead of *The command is started*).

In the English lexicon it is reflected as a fact that many verbs carry the label 'both transitive and intransitive'. The reason of the Agent omission in the «Middle voice» may be that English, while being a nominative type language, has no special means to express the Agent in its ergative model of process, and the transitive means «*by*» is not valid. In Russian the omission of Agent is expressed by the reflective form in the imperfective aspect and in the perfective aspect - by the passive form, while the subject is explicitly marked by the nominative case. In the reflective form for Russian and Bulgarian the Agent still can be freely expressed depending on the participants set. But Czech has some specifics, drawing it nearer to English. In Czech in the reflective form Actor is extremely rare, but due to other reasons. In Czech the reflective form can be used as a realisation of a special type of indicative clause with the general Actor, existing in Slavonic language and not in English, so the Agent in a reflective Czech form is unspecified by definition. In Russian this type of clause is expressed rather by the 3-rd person plural form of a verb. So for English:
The newspapers write (or *It is written in newspapers*)
we have:
Във вестниците пише/ пишат …for Bulgarian,
V novinách se píše - for Czech (other variants are *Noviny píšou, V novinách píšou*)
В газетах пишут - for Russian (other variants are *В газетах пишется, Газеты пишут*).

Another indication of the ergative interpretation in English in opposite to Slavonic languages concerns the possibility that other participants apart from Actor and Goal (for example, Addressee or Beneficiary: *He was given a teapot*) become Subjects in the course of diathesis realisation.

So the four languages—English and three Slavonic ones—under consideration have some forms to express the following three voice meanings. As inflectional languages, the Slavonic exhinbit a more complex situation for passive and middle due to voice – aspect interplay. The following examples are in the present indicative, or in the past indicative (if present is not possible because the perfect aspect meaning is future). We took two sentences for Bulgarian because there is some variation in modern usage - the verb *стартирам* (analogous to the English verb *to start*) is bi-aspectual. The Russian verb *стартовать* with the same stem is bi-aspectual too, but in contrast to the Bulgarian transitive verb it is strictly intransitive; so it is not used in the example under consideration. Thus for Russian and Czech we have only one sentence, with the more traditional verbs *запустить* and *spustít* showing the voice forms.

Active, realised by the verb in active form with the Actee in Direct complement function:

Eng: *The user starts the command; The soldiers build the house.*

Bu, imperfective: *Потребителят стартира командата. Войниците строиха здание.*

Bu, perfective: *Потребителят стартира командата. Войниците построиха здание.*

Cz, imperfective: *Uživatel spouští příkaz.*

Cz, perfective: *Uživatel spustí příkaz.*

Ru, imperfective: *Пользователь запускает команду.*

Ru, perfective: *Пользователь запустит команду.*

Passive, realised by passive form of a verb, Actee in a subject function, Actor in agentive complement function. In Slavonic languages, passive with Actor in the agentive function is used rather in scientific texts. In Czech, periphrastic passive is not very frequent at all; however alternations in the information structure development are prototypically expressed by the word order. Since word order (discussed below) is more flexible in Czech than in the other Slavonic languages under consideration, there is no need to use passive in order to achieve the word order complying to information structure; so the motivations for passivisation are usually different. For instance, it may be used to "conceal" the Actor (either because he is not important or because he is non-specified), to achieve certain text cohesion effects (for instance, to maintain the same Subject), or more generally to maintain parallelism between sentences. In Russian and in Bulgarian, the passive participle form with suffix *-ан* is possible only for perfective aspect, in Czech for both. The passive participle form in imperfective aspect with suffix *-м* is practically not used in the three Slavonic languages and may be ignored in our research. In Bulgarian, there also exist (rather uncommon) passive participles with *-ан* for both types of verbs, e.g.:
Книгата била написана през 1990 г. / Книгата била писана две години.

So:

Eng: The command is started by the user; The house is built by the soldiers;

Bu, imperfective: *Командата се стартира от потребителя. Зданието се строи (строяло) от войниците.*

Bu, perfective: *Командата се стартира от потребителя. Зданието е/било построено от войниците.*

Cz, imperfective: *Příkaz spouští uživatel. (Příkaz je spouštěn uživatelem)*

Cz, perfective: *Příkaz spustí uživatel. (Příkaz je spuštěn uživatelem)*

Ru, imperfective: *Команда запускается пользователем.*

Ru, perfective: *Команда запущена пользователем.*

Middle, characterised by Agent omission. In English it is always expressed by a construction with Subject and active verb form without any Agent expressed with 'by' marker. In Slavonic languages it is expressed by reflective form normally not allowing the presence of an Agent. Often it needs some circumstance to make the clause more natural by content. Bulgarian shows some movement to ergative style : *Командата стартира* as a variant for *Командата се стартира (*see below):

(English) *The command starts; The house builds (by putting blocks).*

(Bu, imperfective) *Командата се стартира*; (*Командата стартира*).

(Bu, perfective) - *Командата се стартира.*

(Bu, imperfective) *Здание се строи.*

(Bu, perfective) **Зданието се построено; Зданието е/било построено или Зданието се строи/ се строяло.*

(Cz, imperfective) *Příkaz se spouští (kliknutím na ikonu);*

(Cz, perfective) *Příkaz se spustí (kliknutím na ikonu);*

I: Příkazy jsou spouštěny dávkově. (Agent odd) - Commands are run in batch-mode.

P: Příkazy jsou spuštěny dávkově.

(Ru, imperfective) *Команда запускается;* (emphasis on the event)

(Ru, perfective) *Команда запустилась;* (emphasis on the result)

The phenomenon of diathesis is also closely correlated with the choice of mood, since both provide realisation of the ideational structure in a speech act. In our case this involves assignment of interpersonal roles for participants (most notably, subject), as well as choices of voice and sensitivity to the textual structure. These choices are also constrained by text style. For example, instructions can be expressed as imperative or infinitive clauses (Actor is omitted, Actee is a Direct-Complement), impersonal suggestions (Actor is omitted, Actee is a Subject), modal impersonal suggestions (Actor is omitted, Actee is a Direct-Complement, modality is expressed by a predication, *можно*, in Russian), invitations (Actor is expressed by the first person pronoun, either singular or plural, Actee is a Direct-Complement). These topics will be addressed in the final demonstrator.

2.3 Mood

The system of MOOD, as understood in Hallidayan terms, belongs to the interpersonal metafunction of the language and is the grammatical resource for realising a specific speech function, i.e., an interactive move in dialogue (Martin et al., 1997). The alternative MOOD choices, i.e., indicative, imperative, interrogative, are associated with different interaction status. The basic MOOD options at the level of the clause are imperative and indicative; the latter subdivided into interrogative and declarative mood.

The distinction between the imperative and the indicative mood holds for all three Slavonic languages under study here, Bulgarian, Czech, and Russian, and is manifest in the three corpora. The interrogative mood, which is also normally valid as an option in these languages, has no occurrences in the corpora due to the constraints of the genre. In terms of distribution of the different mood options in the corpus, roughly the same number of occurrences of the imperative and the indicative (declarative) are to be found, which is characteristic of this specific text genre.

The indicative mood employs verb forms in the respective tense and aspect options marked for the grammatical categories relevant in a particular language. The imperative mood employs specific imperative verb forms which are marked for the category of number, i.e., singular vs. plural, but not gender or tense, for example. Such forms are predominantly in the second person in all three languages, although there are exceptions such as the equivalent of the English suggestive «let's» type. The singular vs. plural distinction of verb

forms in the imperative mood serves two functions: (a) referring to a number of addressees, more than one; and (b) indicating politeness, formality, stylistic differences, or register, of the discourse. In the corpora of the three languages, this distinction in number refers to the second dimension above, i.e., politeness and formal register.

Halliday makes a distinction between (a) MOOD as the interpersonal system in the grammar and (b) Mood as the interpersonal element of the clause that consists of Subject and Finite. The Mood in English consists of the combination of the Subject and the Finite, which can be verified by tests such as agreement, tagging, and ellipsis/substitution. The Mood element of the three Slavonic languages here includes the Finite (a verb form) and possibly the Subject, e.g., as elicited by the agreement test and the ellipsis/substitution test for identification; the tagging test is not applicable in these languages. The three languages may be found to differ on the point of inclusion of a number of verb forms in the Mood (Finite) element due to the considerable differences in the verbal tense systems and the composition of the verb phrase. The Mood element typically consists of the Subject and the Finite but all three are pro-drop languages, thus permitting omission of the Subject, or, in other words, allowing for non-explication of the Subject.

Thus, the Mood element may consist of the Finite only, structurally speaking, on most occasions, although it is also in agreement with an implicit subject in terms of the categories of number, gender, etc. in the respective language. In addition, the Subject-Finite ordering is not a defining feature, or means of realisation, for mood, i.e., the distinction of the interrogative vs. the declarative is not tied to a difference of Subject-Finite ordering. This is so because of the relatively 'free,' or flexible, word order in the three languages. Instead, suprasegmental information is frequently used to distinguish between moods, as well question words and question particles such as *ли* in Russian and Czech.

The Mood element in the clause is also the locus of realisation of the interpersonal resources of POLARITY and MODALITY, apart from MOOD. In all three corpora there are very few examples of negative polarity, e.g., in Czech,

Příkaz ODRUKY nepřijímá souřadnicový vstup. (does not accept)

As for modality, just as MOOD, it is treated as a separate lexicogrammatical feature. It has the following basic options: notmodal, probability, usuality, inclination, ability, and obligation. Modality is realised, or can be expressed, as a feature of the Finite or by a separate mood Adjunct. As a feature of the Finite, Modality is realised by means of modal auxiliaries such as *can, could, must, should*, etc. in English, for example. The corpora in the three Slavonic languages contain occurrences of modality realised by means of modal auxiliaries (in the Finite) such as *искам* (7 occurrences of the volitional type), *мога* (9 occurrences of the possibility type), and *трябва* (1 occurrence of the necessity type) in Bulgarian; *можно* (a so-called impersonal, or non-finite modal verb) and *позволяет* (finite verb) in Russian both referring to the modal option of possibility. In Czech, such auxiliaries are *moci* (ability), *muset* (obligation), *nemuset* (needn't), *chtít* (inclination), *něco má být* (something should be) = *mít* (have) in modal use.

Modality in the three languages can also be realised by mood Adjuncts, e.g., adverbs such as *probably, perhaps, readily, willingly*, etc. in English, and their equivalent in Slavonic languages, for example, *вероятно, може би, охотно*, etc. in Bulgarian and *возможно, может быть*, etc. in Russian, and *opravdu* (really), *jistě* (surely) *snad* (perhaps), *obvykle* (usually), *pravděpodobně* (probably), *ochotně* (willingly), *nechtěně* (unintentionally, «unwantedly») in Czech. Due to the characteristics of the genre, in only one of the three

corpora, however, modality has been found to be realised by mood Adjuncts, i.e., in Czech – *raději (rather, preferably).*

2.3.1 Description for Bulgarian

2.3.1.1 Theoretical background

Theoretical grammars of Bulgarian differ on their interpretation of the above phenomena. Modality is rarely, if ever, discussed as such. Most authors concern themselves with mood only, and particularly, the realisation of mood in terms of specific wordforms or combinations of wordforms.

The traditional treatment of mood in Bulgarian distinguishes among the following options: indicative, imperative, conditional, and non-evidential (Stojanov, 1980) on the basis of the different attitudes of the speaker to the action expressed by the verb, or rather, as Pashov (1989) points out, the different kinds of relationship of the action to reality, i.e., real, or possible, desirable, imperative, etc. A more satisfactory account would be to interpret the above mentioned «moods» in terms of the speaker's belief of the kind of relationship that holds between the action and reality, i.e., the speaker's belief that the action is possible, or desirable, etc. to take place in reality.

On traditional views, the four moods in Bulgarian are characterised thus:

1) in Indicative, the action is presented as real, i.e., it indicates the presence or fact/reality of this action (in the past, present, or future);
2) in Imperative, it expresses an urge on behalf of the speaker to the hearer, who is also the verb's subject, to carry out the action referred to;
3) in Conditional, the speaker refers to possible (under certain conditions), imaginary, or desirable actions;
4) in Non-evidential, the speaker does the following:
    (a) presents oneself as a non-witness or non-participant in the action
    (b) expresses doubt, suspicion, mistrust, or disagreement with someone's opinion
    (c) expresses surprise in reference to the action described by the main verb.

The fourth mood above, the non-evidential mood, is the most controversial one in theoretical discussions of Bulgarian. Numerous varying accounts of the mood system diverge on this issue. Treatments range from considering it a separate mood, through a kind of «supermood,» to an explanation of verb-forms not belonging to the category of mood, but a special morphological category.

2.3.1.2 Evidence from the corpus

Theoretical disputes apart, the analysis of the Bulgarian corpus elicits the following conclusions concerning mood. The distinction between indicative and imperative holds here as well with the two types accounting for roughly the same number of clauses. This is, inevitably, an inherent bias in this sort of text/discourse whose main purpose is instructional. In other texts, the imperative mood is much less frequent. At the same, there is not a single occurrence of the interrogative mood option, which is also non-representative of the language as a whole. Due to the distinction between the so-called 'polite' and 'impolite', or rather, 'formal' and 'informal' forms of address used in Bulgarian as in other Slavonic languages, the imperative forms can be either in the 2nd person Singular or in the 2nd person Plural. The translation in the corpus employs the 'formal' or 'polite' forms of address; thus all imperative verb forms are in the 2nd person Plural.

As regards modality, in the Bulgarian corpus there are no occurrences of modality expressed by mood adjuncts. Modality is only realised, even if rarely, by means of modal auxiliaries (the Finite) such as the volitional *искам* (7 occurrences), the possibility one *мога* (9 occurrences), and the necessity one *трябва* (1 occurrence).

2.3.2 Description for Russian

2.3.2.1 Theoretical background

The grammatical indicative mood is a standard way in Russian (as well as other languages) to deliver information about the state of affairs in the form of a proposition. In the indicative mood, the Subject is marked by the nominative case, and agrees with the Finite in number and person in the present/future tense and in number and gender in the past tense. Two other mood forms in Russian are imperative and conditional. The most prevalent form in our corpus is imperative, so we'll consider it in grater detail. The two basic imperative forms are:

- imperative proper – *Укажи(те) конечную точку (Specify the endpoint);* the imperative form expresses a request or an order directed to the Interactant-another excluding the Speaker; the plural finite expresses the politeness;
- suggestive *–Давай(те) укажем конечную точку (Let's specify the endpoint);* an invitation to an action directed to the Interactant-another including the speaker. It is expressed by the first-person plural finite form, or by the auxiliary verb *давай(те)* (the imperative form of *давать*) + the first-person plural finite form of the predicator. Politeness is expressed by the ending *-те*.

When in semantics we have some relation allowing modal realisation, for example, a habitual function, destination of a computer system concept, in English it is represented by Subject - process relation (OFFSET create), but in Russian often it is better to realise this meaning of habituation, possibility by modal words. In this case we can use active constructions in impersonal or personal form of the sentence:

impersonal: Можно указать эти значения [58] (One can enter these values)

impersonal: Командой OFFSET можно создать копии линий (One can create copies of lines by the OFFSET command); the nominal group in the instrumental case conflates the semantic roles of Instrument and Actor.

personal active: Командой OFFSET может (позволяет) создавать копии линий (OFFSET can (allow) create copies of lines);

personal quasi-passive: Командой OFFSET могут создаваться копии линий (You can create copies of lines by the OFFSET command).

2.3.2.2 Evidence from the corpus

The following ways are possible for expression of instructions for use in Russian: clauses in imperative mood (*Нажмите кнопку*), infinitive clauses (*Нажать кнопку*), impersonal indicatives (*...нажимается кнопка*) and impersonal modal assertions (*...можно нажать кнопку*). Imperative expressions are standard in the register of manuals translated from English, while original Russian manuals often tend to use infinitive or impersonal constructions. Since our corpus is translated from the English CAD manual, it continues the English tradition in use of the imperative mood.

Elaborations in instructions are expressed in the indicative mood:

Команда LINETYPE задает тип линии.
LINETYPE sets the current linetype.

2.3.3 Description for Czech

2.3.3.1 Evidence from the corpus

The description for Czech is not radically different from that for Bulgarian and Russian. In the corpus, there are 18 occurrences of modality realisations, including 3 of inclination, 11 of ability, and 4 of obligation. Here are some examples of occurrences of explicit modality in the Czech corpus:

1. *Novou úsečku můžete začít v koncovém bodu (6) poslední nakreslené úsečky stisknutím ENTER při dotazu na Počáteční bod.*
   You can start a new line in the end point of the last drawn line by pressing ENTER at the Start point prompt.
2. *Můžete zadat tyto relativní hodnoty ve formě @vzdálenost<úhel*
   You can enter these relative values in the format @distance<angle
3. *(v tomto případě, raději zadejte @3<100).*
   (In this case, enter rather @3<100)
4. *Poté co vytvoříte křivku, ji můžete editovat pomocí příkazu KEDIT*
   After you create a polyline you can edit it using the KEDIT command.
5. *nebo použít příkaz ROZLOŽ pro přeměnění křivky na jednotlivé úsečky a obloukové segmenty.*
   or use the ROZLOŽ command to convert the polyline into individual line and arc segments
6. *Element nemusí být nakreslen v počátku.*
   The element need not be drawn at the origin.
7. *Každý element může mít různou barvu a typ čáry.*
   Each element can have a different color and line type.
8. *Pod Popis zadejte popis (nepovinné), který může mít až 255 znaků včetně mezer.*
   Under Description enter a Description (optional), which may have up to 255 characters including spaces.
9. *Poté co jste vytvořili polygon, můžete jej editovat pomocí příkazu KEDIT*
   After you created a polygon, you can edit it using the KEDIT command.
10. *nebo jej zkonvertovat na jednotlivé úsečkové segmenty příkazem ROZLOŽ.*
    or convert it into individual line segments using the ROZLOŽ command
11. *Stiskněte znova výběrové tlačítko a pero se zdvihne tak, že nyní můžete pohybovat kurzorem po obrazovce bez kreslení.*
    Press the choice button again and the pen will lift so that you can move the cursor on the screen without drawing.
12. *Jestliže je pero dole, můžete pokračovat v kreslení po provedeném záznamu.*
    If the pen is down, you can continue drawing after recording.

13. *Pokud budete chtít používat režimy Krok {během skicování} a Orto během skicování, musíte je přepínat z klávesnice pomocí kláves F8 pro Orto a F9 pro Krok.*
    If you want to use the Step and Orto mode (during free hand sketching) and you have to switch between them from the keybord using the F8 key for Orto and F9 for Step.

2.3.4 Formal specifications

The basic SFG MOOD system has to express the following clause classification [indicative, imperative] that is natural for Slavonic languages. There is one main difference comparing English with Bulgarian/ Russian/ Czech imperative clause. In Slavonic languages finite verb form in 2nd person is used for their imperatives versus nonfinite (infinitive) form of the verb in English imperative sentences. The following formalization describes the Slavonic Mood-Type and inserts Finite for imperative clauses:

```
MOOD-TYPE
Independent-clause-simplex->
    [indicative]
    [imperative] (+Finite)
```

2.3.4.1 Indicative Clauses

The three Slavonic languages use the same model of indicative clauses:

```
INDICATIVE-TYPE
indicative->
    [declarative]
    [interrogative]
```

As shown by the theory, the main task dealing with indicative clauses is the determination of the MOOD element (roughly we can say «Subject – Finite pair»). When the MOOD element is once inserted by the system FINITE-CLAUSE it starts collecting its features from the dependent systems so that the combinations of different features gives the MOOD element different grammatical (morphological) forms which represent different nuances in respect to semantic meaning of the clause. Important notions are modality, tense, and aspect, which are represented below separately as far as their examples occur in the chosen corpus.

All indicative clauses in our corpus are declarative because of the nature and style of instructional texts.

2.3.4.2 Imperative Clauses

2.3.4.2.1 Formal specifications for Bulgarian

Dealing with the imperative clauses one needs to infer the proper finite verb form. The system IMPERATIVE-INTERACTANT-SUBJECT when the output [jussive] is chosen is the right place to constrain the finite verb at the imperative clause to its 2nd person form:

```
IMPERATIVE-INTERACTANT-SUBJECT
imperative-interactant->
[jussive] (Finite:::secondperson-form)
[oblative] (Subject ! letmepronoun)
[suggestive] (Subject ! letuspronoun)
```

The adequate reflection of the imperative verb form in Bulgarian needs a new system in the grammar net used to capture the difference between polite and personal imperative, or rather formal/informal. The morphological marker for politeness is plurality, so the system could be defined on the base of the following informational rule:

```
IMPERATIVE-MOOD-TYPE
jussive->
[polite-imperative] (Finite:::plural-form)
[personal-imperative] (Finite:::singular-form)
```

The semantic difference between imperative sentences addressed to the group of people and to one person in polite form should be underlined:

```
POLITE-IMPERATIVE-NUMBER
polite-imperative->
    [person-addressee]
    [group-addressee]
```

All imperative clauses in the corpus have no explicit subject, although the subject could be used in imperatives of the three languages. Both polite (formal) and personal (informal) use of the imperative mood can have implicit or explicit subject.

```
IMPERATIVE-SUBJECT-PRESUMPTION
imperative->
[imperative-subject-implicit]
[imperative-subject-explicit]  (Subject : Addressee)
```

2.3.4.2.2 Formal specifications for Czech

The following specifications express Czech imperatives with their grammatical features specific for the language:

```
IMPERATIVE-INTERACTANT-SUBJECT
imperative-interactant->
    [jussive]            (Finite:::secondperson-form)
       (Finite:::imperative-form)
    [suggestive]   (Finite:::firstperson-form)
       (Finite:::plural-form)
       (Finite:::imperative-form)
    [oblative]     (+Conjunct)
       (Conjunct ! at3)
       (Finite:::indicative-form)

IMPERATIVE-MOOD-TYPE
jussive->
    [polite-imperative] (Finite:::plural-form)
    [personal-imperative] (Finite:::singular-form)
```

```
POLITE-IMPERATIVE-NUMBER
polite-imperative->
    [person-addressee]
    [group-addressee]

IMPERATIVE-SUBJECT-PRESUMPTION
imperative->
    [imperative-subject-implicit]
    [imperative-subject-explicit]  (Subject : Addressee)
    (Subject:::Personal-Pronoun)
```

2.3.4.2.3 Formal specifications for Russian

The following specifications encode different types of imperatives in Russian:

```
IMPERATIVE-MOOD-TYPE
imperative-interactant->
   [jussive] (Finite:::imperative-form)
   [suggestive] (Finite:::plural-form) (Finite:::1-person-form)

IMPERATIVE-POLITENESS-TYPE
(jussive) ->
   [polite-imperative] (Finite:::plural-form)
   [personal-imperative] (Finite:::singular-form)

SUGGESTIVE-MOOD-TYPE
(suggestive) ->
   [invitation] (+AuxDavatj:::imperative-form)
   [suggestion]
```

The Subject-Verb agreement in the indicative mood is achieved in the same way as in the English grammar.

2.4 Tense

2.4.1 Theoretical background

The grammatical category of Verb Tense refers the action described by the verb to the temporal line in relation to the time of speaking, i.e., the main reference point, or, alternatively, simultaneously to the main point and to a secondary reference point.

The three languages under investigation here differ on the extent of elaboration of their respective verb-tense systems, e.g. Bulgarian has a rather complex system with a considerable number of tenses (at least 9) whereas the Russian and Czech systems encompass 3 tenses only—past, present, and future. These two languages differ on the number of reference points used, e.g., for Bulgarian (and English) both main and secondary reference points are relevant whereas in Russian, orientation to one temporal point is used only.

Most generally, verb tenses in all three languages, including the rather more complex case of Bulgarian, can be classified into the following three subcategories:

(a) the action is simultaneous with the moment of speaking—present tenses;
(b) the action precedes the moment of speaking—past tenses;
(c) the action follows after the moment of speaking—future tenses.

As already mentioned, at the same time, verb tenses in Bulgarian may have a simple (1-point) or a complex (2-point) temporal orientation, e.g. *четях* «I was reading» where reference is made to a second orientation point in the past (in addition to the moment of speaking). In this way, the number of verb tenses in Bulgarian is considered to be 9, i.e.:

1. The Present Tense
2. The Past Perfective Tense
3. The Past Imperfective Tense
4. The Past Indefinite Tense
5. The Past Perfect Tense
6. The Future Tense
7. The Future Perfect Tense
8. The Future in the Past Tense
9. The Future Perfect in the Past Tense.

In all three languages, the past tense (the Past Indefinite in Bulgarian) is formed by means of the so-called L-participle or L-form, i.e., a form of the verb inflected by «-l», «-la», «-lo», «-li» to signal past tense but also gender and number of the referent with which/who the action of the main verb is associated. The difference lies in the fact that this form is sufficient to express past tense in Russian whereas in Czech and in the Bulgarian Past Indefinite, for example, the indicative present form of the auxiliary verb *být* and *съм* (to be), respectively, plus the past participle (the L-participle) are used.

The present tense of imperfective verbs is formed by the indicative present form. The simple future tense in Bulgarian is formed by means of a combination of the 'future' particle *ще* and the conjugated (finite) form of the verb, i.e., analytical forms; this applies to both perfective and imperfective verbs. The future tense of imperfective verbs is formed by the future form of the copula *být* (in Czech) and *быть* (in Russian) and the infinitive, i.e., by means of analytical forms again. However, the future tense of perfective verbs is created by the indicative present form of the verb. There are several (imperfective) verbs in Czech and Russian that have simple future forms, i.e., non-analytical. These are verbs of movement and the verbs *být* in Czech and *быть* in Russian. The following table illustrates the forms of the 1st person singular masculine of various verbs in all tenses in Czech, for example.

	past	present	future
Číst, imperf. (to read)	*četl jsem*	*čtu*	*budu číst*
Přečíst, perf. (to read over)	*přečetl jsem*	—	*přečtu*
Jet, imperf. (to ride)	*jel jsem*	*jedu*	*pojedu*
Být, imperf. (to be)	*byl jsem*	*jsem*	*budu*

Generally speaking, the indicative verb forms express three tenses in Czech and Russian and are used in all tenses in Bulgarian. This statement, however, is fully true for imperfective verbs only. Perfective verbs, by their nature, cannot express the present. From the morphological point of view, however, they have two forms—past and present—and the present form has meaning of future.

Furthermore, in all three languages, the clauses in the imperative mood do not distinguish tenses but only aspects which is related to the treatment of aspect as a lexico-grammatical rather than a purely morphological (grammatical) category.

In modern Russian the tense meanings relate the Finite element of a clause directly to the moment of speech as simultaneous to it (the present tense), previous (the past tense) or next (the future tense). In gerund and participial forms the tense of a process is relative to the tense of the Finite. Tense is closely related to aspect. The full paradigm (past - present - future) is available only for verbs in the imperfective aspect -З*апускал - запускаю - буду запускать программу - I started - start - will start the program*; for verbs in the perfective aspect there are only two tense forms (past - future), for example, *запустил - запущу программу - I started - will start a program.* Concrete meaning of a tense form is greatly defined by the aspectual meaning of the verb (see the section on Aspect).

2.4.2 Evidence from the corpus

In the corpora used in the project, only a limited number of tenses occur. Thus, for the Bulgarian, only two out of nine tenses are attested—the Present Tense and the Past Indefinite Tense—whereas only one tense is to be found in the Russian corpus, namely the Past Tense. In the Czech corpus, most sentences are in the future or the past tense. The present tense occurs only rarely in the Czech corpus.

Interestingly, the distribution of tenses in the different types of clauses, as distinguished by MOOD, varies. In the corpora, tenses vary in the clauses in the Indicative mood which function descriptive and/or instructive purposes.

In the Bulgarian corpus, the Present Tense is the predominant verb tense in the corpus, occurring both in main clauses,

e.g., (86-7) *Линиите пресичат края на цялата мултилиния, а външните дъги съединяват краищата на най-външните елементи*

which corresponds to the English page 48 *Lines cross the end of the whole multiline and outer arcs join the ends of the outermost elements*, and in subordinate clauses,

e.g., (2-3) *Стартирайте командата LINE, като използвате един от следните методи,*

which corresponds to the English *Start the LINE command using one of these methods.*

The Past Indefinite Tense in Bulgarian occurs several times in the corpus, much less frequently than the Present Tense. It is roughly equivalent to the Present Perfect (or sometimes Simple Tenses in English) and its use can be illustrated by the following example:

(227-8) *След като вече сте записали линиите, начертани на ръка, не можете да ги редактирате,*

which corresponds to the English (page 53) *Once you record freehand lines, you can't edit them*. Thus, out of the rather elaborate system of 9 tenses, only 2 occur in the limited scope of the corpus, one referring to a «present» moment, and another referring to a point in time preceding the «present/current» point, the latter much less frequently than the Present Tense.

In the Czech corpus, as already mentioned, most sentences are in the future or the past tense with rare occurrences of the present tense. The Future tense is used for the description of what happens when the user performs a certain action, or what will be contents of the next example. The Present tense is used for the description of objects that have been already created, and for modal verbs. The Past tense is used for the description of objects that have been created and the next action is a continuation of the whole process.

The following are the occurrences in the Czech corpus (the translation to English is word-for-word so that the tenses could be understood well):

1. *V dalším příkladu si nakreslíte*(fut.) *rovný segment křivky, na který navazuje*(pres.) *obloukový segment, pak nakreslíte*(fut.) *jiný rovný úsek v tečném směru.*
   In the next example you will draw a straight segment of the curve, to which the arch segment is tied, then you will draw another straight segment ...
2. *Budete definovat*(fut.) *poloměr, úhel vyplnění a koncový bod oblouku.*
   You will define radius, angle and the end point of the arch.
3. *Stiskněte opět tlačítko a kreslení bude obnoveno*(fut.) *od nové polohy kurzoru.*
   Press the button again and the drawing will continue from the new position of the cursor.
4. *Pokud budete chtít používat*(fut.) *režimy Krok {během skicování} a Orto během skicování, musíte*(pres.) *je přepínat z klávesnice pomocí kláves F8 pro Orto a F9 pro Krok.*
   In case that you will want to use the modes Step and Orto during the sketching, you have to switch them on the keyboard ...
5. *Poté co jste vytvořili*(past) *polygon, můžete*(pres.) *jej editovat pomocí příkazu KEDIT*
   After you created polygon you are able to edit it with the command KEDIT
6. *Pohybujte kurzorem na konec čáry, kterou jste nakreslili*(past) *naposledy*
   Move the cursor to the end of the line that you drew last.
7. *Při rozkládání široké křivky se šířka čáry změní*(fut.) *na 0 a výsledná křivka je umístěna*(pres.) *podél středu, kde byla*(past) *křivka.*
   During decomposition of a wide curve the width of the line will change to 0 and the resulting curve is placed along the center, where the curve was.
8. *Jestliže bylo*(past) *pero dole, zdvihne se*(fut.) *nahoru.*
   If the pen was down it will raise up.
9. *Aktuální pohled můžete*(pres.) *měnit během skicování, jen pokud je*(pres.) *pero nahoře, všechny čáry byly již zaznamenány*(past) *a režim tabletu je vypnut*(pres.).
   You are able to change the current view during the sketching, only if the pen is up, all lines were already recorded and the mode of the tablet is switched off.

Tense in imperative clauses, which are the most frequent case in our corpus, is unspecified in Russian, so temporal forms of predicates are used in impersonal constructions (*эти данные **указываются** в следующей форме*) and elaborations (*команда **задает** тип линии*), which are given in the present tense. The only occurrence of the past tense (in the passive form) and the future perfective tense (in the reflective form) happens in one sentence: *Если перо было опущено, оно автоматически поднимется - If the pen was down, it moves up.*

2.4.3 Formal specifications

2.4.3.1 Formal specifications for Russian

```
TENSE-SYSTEM
(clause-simplex) →
   [past] (Finite:::past)
   [future]
   [present] (Finite:::present)

FUTURE-TENSE
(future) →
   [perfective] (Finite:::present)
   [imperfective] (+AuxBe/Finite) (Lexverb:::infinitive)
```

2.4.3.2 Formal specifications for Bulgarian

```
TENSE-SYSTEM
(clause-simplex) →
   [past]
   [future] (+Ste-aux ! Ste) (Finite:::present-form)(Ste-aux^Finite)
   [present] (Finite:::present-form)

PAST-TENSE
(past)->
   [past-simple] (Finite:::past-form)
   [past-perfective] (Finite:::present-form) (Finite:: sum-aux)
(+Past-dependent) (Past-dependent::lparticiple) (Finite^Past-dependent)
```

2.4.3.3 Formal specifications for Czech

```
TENSE-SYSTEM
(clause-simplex) ->
   [past]   (+AuxBe/Finite)
      (Finite:::present)
      (Lexverb:::pastparticiple)
   [future]
   [present]      (Finite:::present)

FUTURE-TENSE
(future) ->
   [perfective]          (Finite:::present)
   [imperfective] (+Copula/Finite)
            (Finite:::future)
            (Lexverb:::infinitive)
```

2.5 Aspect

2.5.1 Theoretical background

The system of tenses in Slavonic languages is closely interrelated with the category of aspect to the extent that it is often referred to as the tense-aspect system of the verb.

Aspect is a characteristic feature of Slavonic languages, a grammatical category belonging to the verbal system. The category of aspect in Slavonic languages is related to the way the nature of the action is described, in particular the action's temporal characteristics such as its duration, frequency of occurrence, completion, etc. These temporal characteristics, however, are not directly associated with the predominantly deictic reference to temporality,

e.g., past, present, and future time periods, which is typically delegated to the domain of the verbal category of tenses.

In Hallidayan theory, aspects appear to be a function of the Predicator (which is itself part of the Residue), together with secondary tense, voice, and type of process. It is important to note, however, that mainstream Hallidayan linguistics has not contributed extensively to the study of aspect, probably because of the nature of the language systems which it is predominantly based on.

In Slavonic languages, a distinction is made between two aspects—the perfective and the imperfective, although it is difficult to claim that scholarly consensus has been reached on the exact definition of the two types and their forms as some authors prefer to distinguish between resultative and non-resultative aspects of the action, for example in Bulgarian. The perfective and imperfective verbs form pairs with the same meaning and valency frames.

Thus, in the Slavonic languages each verb belongs to (at least) one of the two classes, imperfective and perfective. Generally speaking, the meaning of the two aspects is distinguished in the following ways. The imperfective verb forms emphasize the following semantic features in the description of the action:
(a) the duration and continuous, uninterrupted nature of the activity;
(b) the incomplete nature of the activity;
(c) the repeated, recurrent, habitual, or otherwise frequentative nature of the activity;
general statements/truths.

The perfective aspectual verb forms exhibit the following set of features:
(a) they refer to the action as a whole;
(b) they focus on the completion of the action.

A specific characteristic of the category of verb aspect in Slavonic languages is its lexicalization, the regularity of occurrence of aspectual pairs which can be formed by a combination of a perfective and an imperfective verb. This consideration, among others, leads to the conclusion that in Slavonic languages verb aspect is a lexico-grammatical and not a merely morphological feature.

In this binary category of aspect, the imperfective forms are the unmarked ones in terms of range of usage, default value, frequency of occurrence, degree of collocability. For example, perfective verbs are not used in combination with phase verbs such as *начать* («begin»), *кончить* («finish»), *продолжать* («continue») in Russian and the other Slavonic languages; negative imperative clauses do not employ perfective verbs, etc.

The unmarked nature of the imperfective verb forms is evident also in derivation patterns–usually the perfective form is derived from the imperfective one. Thus, in the aspectual pair, verb forms of one aspect can be derived from the respective forms of the other aspect. It is even possible to form a network of semantically and morphologically related verb forms, e.g., basic imperfective verb —> perfective verb derived by prefixation —> imperfective verb created by changing the stem-forming affix —> perfective verb derived by prefixation. An example of such a «chain-like» network is the verb *hnát* (to drive cattle, to make sb run somewhere) in Czech and its derivatives:
hnát (to drive)
zahnat (to drive away, perfective)
zahánìt (to drive away, progressive)
pozahánìt (to drive away all, one after another, perfective).

The derivation of aspect counterparts, however, shows many irregularities and exceptions. There are usually more than one prefix that can be attached to the basic or primary verb and some of them may substantially change the meaning and/or the frame of the verb, e.g. *писать—переписать*, *записать, выписать*, etc. in Russian and *nést* (to carry), *pøinést* (to bring), *odnést* (to take away), *snést* (to bear a situation, to lay an egg), *vznést se* (to fly up) in Czech. Furthermore, some verbs do not enter aspectual pairs and are either perfective only or imperfective only. In addition, there is a growing class of verbs which exhibit dual aspectual characteristics, especially loan words, e.g., *стартирам* in the Bulgarian corpus.

If we are looking for aspectual pairs, we cannot simply rely on the derivational means mentioned earlier. Prefixation or change of suffix is not always possible, the verb may have prefixes that cannot be removed, or the verb without a prefix may already be perfective. There are also aspectual pairs consisting of different roots, e.g., *brát*-imperf.—*vzít*-perf. (to take) in Czech and their cognate pairs in Bulgarian and Russian. As we mentioned in the previous paragraph, some verbs do not have any aspectual counterpart, they exist only in the perfective or imperfective meaning, or they can be understood as both perfective and imperfective. Here are several examples from Czech:

- imperfective only, e.g. *propagovat* (to propagandize), *bőt se* (to be afraid),
- perfective only, e.g. *pouит si (co)* in *Co si pouneme?* (What shall we do?),

dual aspect verbs (one form for both aspects), mostly loan words ending with *-ovat*, e.g. *identifikovat* (to identify), but also several original Czech verbs, e.g. *soustшedit se* (to concentrate).

Thus, in general, there are several types of verbs according to their association or lack thereof with a specific aspect. These fall into the following broadly defined classes:
(a) imperfective only, i.e., lacking a perfective counterpart;
(b) perfective only, i.e., lacking an imperfective counterpart;
(c) dual aspect verbs, i.e., both aspects can be expressed by means of the same forms;
(d) verbs belonging to aspectual pairs, i.e., two different forms with a aspectual meaning—these being the norm in our languages.

The distribution and usage of aspectual verb forms with respect to MOOD and TENSE. In general, the imperative mood employs verbs of both types of aspect, perfective and imperfective. In negative imperative clauses, however, mostly verb forms of the imperfective aspect are used, e.g., *„Zavøete okna!«*, *„Nezavírejte okna!«* (Don't close the windows!) in Czech, or similarly, «*Отворете/Отваряйте прозореца*!» (Open_perf./Open_imperf. the window!) vs. *«Не отваряйте прозореца!»* (Don't open_imperf. the window!) in Bulgarian.

In the indicative mood, the usage of aspect depends on the nature of the described process or action. For progressive, iterative and frequentative processes, the imperfective aspect is used while for completed processes it is the perfective aspect. However, these patterns interact significantly with the usage of different tenses as well, as we already mentioned above.

In Russian the imperfective aspect has the following basic meanings:

- continuous, describing an event in its middle phase, which beginning and ending are not foregrounded, for example, *Вечером я писал письмо - In the evening I was writing a letter*;

- factual-mentioning, in which only a fact of an event is mentioned without reference to its output event: *Команда LINETYPE задает тип линии (25) - LINETYPE sets the current linetype* (in contrast to *Тип линии, заданный командой LINETYPE, ...– The linetype which has been set by LINETYPE...*, which is expressed using the perfective verb);
- repeated, in which an action denoted by a verb is treated as routinely repeated. This case is present in our corpus: *Нажимайте клавишу r каждый раз... (193) - Enter r at any time to record.*

The basic meaning of the perfective aspect in Russian is factual, which refers to an event as a complete whole: *После того как прямоугольник нарисован ...(174) - After you've created a polygon ...* Other meanings (exemplification and potentiality) are absent in the register.

Though in Russian aspect and tense refer to different facets of time reference; some of perfective meanings are related to the Russian tense system: the future tense for the imperfective verbs is produced using the auxiliary verb *быть* (to be) followed by the infinitive, while the semantic present tense is impossible for the perfective verbs (their morphological present tense means only the future). This is justified by the fact that the perfective aspect being used in the indicative mood emphasises the completion of an event, which either has been already happened or will happen. In the imperative clauses the imperfective aspect means repetition of an action, while the imperfective aspect—the action in its completion. Thus, the overwhelming majority of imperative clauses in our corpus use verbs in the perfective aspect. In other styles possible in the same register, commands can also be expressed in infinitive clauses or impersonal modal assertions still using the perfective verbs (cf. discussion on the mood in Russian). Suggestions for actions or comments are typically expressed using verbs in the imperfective aspect (*Поле вставляется нажатием клавиш CTRL-F9 - The field is inserted by pressing CTRL-F9; Команда LINETYPE задает тип линии - LINETYPE sets the current linetype*), since they do not emphasise the completion of action, rather its possibility. Negations in imperative clauses can mean:

- in the imperfective aspect: prohibitions or warnings; this is not present in our corpus, but possible in the register: *Не запускайте* (**запустите*) *команду, пока не убедитесь ...-Do not start the command until smth is ensured*;
- in the perfective aspect: personal suggestions or cautions; this is impossible in this register (mostly used in personal communication: *Не упади-Do not fall*).

2.5.2 Evidence from the corpus

In the corpora under investigation here, the imperative mood clauses employ mostly perfective verb forms but also some imperfective ones, as an exception, at least in Bulgarian and Russian. In the indicative mood, both types are used but mostly imperfective verb forms which is related to the descriptive intent of this type of instructional text, especially one based on the description of a procedural schema.

In the Bulgarian corpus, verbs of both aspects occur, with some biases which are rather idiosyncratic of the corpus and not necessarily representative of the language as a whole, i.e., the prevalence of perfective verb forms. Out of the two moods, imperative, and indicative, the former is predominantly populated by perfective verbs, e.g., *преместете,*

въведете, натиснете, with very few exceptions, e.g., (191) *въвеждайте* ***r*** *периодично,* which corresponds to the English, *Press* ***r*** *any time* (page 52), and the latter is mainly realised by imperfective verbs, e.g., *като използвате*. There are also verbs of dual aspect, e.g., *стартирайте* in the Bulgarian corpus.

In AGILE corpus the perfective aspect in the imperative form occurs in the overwhelming majority of Russian clauses. The only exception is one occurrence of repeated meaning in the imperfective aspect: *Нажимайте клавишу r каждый раз... (193) - Enter r at any time to record ...* Explanations, which are expressed in the indicative mood, use the imperfective aspect:

Команда LINETYPE задает тип линии (25) - LINETYPE sets the current linetype;

vs.

Задайте тип линии командой LINETYPE
Set-perf *type line-gen command*-inst *LINETYPE.*
Set the line type by the LINETYPE command.

Side effects are expressed in the perfective aspect, since they describe completion of the process:

На экране появится диалоговое окно AutoCAD Text Window. [215]
On screen-loc appears dialogue box AutoCAD Text Window.
The AutoCAD Text Window appears.

Aspectual pairs exist for all verbs in the corpus except some relational processes (*появиться, состоять*).

2.5.3 Formal specifications

2.5.3.1 Bulgarian

As far as in Bulgarian language the idea of ASPECT is carried by the verbs, which have different lexical presentation for one and the same process in perfective and imperfective aspect, each verb as lexical item should contain in the list of its features an appropriate marker/s.

- for imperfectivness: "incomplete-process", "repeated-action"
- for perfectivenesss: "complete-process", "terminative-action"

Grammatical resources should provide the choice of the lexical item, when the meaning of the clause is based on the verb ASPECT. For example :

Въведете и (16)
Enter u

or

Въвеждайте r периодично (191)
Enter r at any time

Both phrases are imperative, in 2nd person plural. When the English variant is generated there are no reasons to think about such choice of the verb- the lexical item is one and the same in both cases. The idea of ASPECT in the second clause above is realised trough the adverb *at any time.* Although the correspondence of that adverb exists in Bulgarian

translation *периодично* the notion of «incomplete nature of the process» is carried by the verb *въвеждам* and logicaly this lexical item has as its feature "repeated-action".

ASPECT is a characteristic of clause and refers to the Predicator, so the initial ASPECT system should be located at clause rank:

```
ASPECT
clause-simplex->
    [perfective] (Process::perfective)
    [imperfective] (Process::imperfective)
```

That marker of the Process (perfective/imperfective) will be used in the Grammar net later, where are there some differences in use and form of the two verb types. For example, in Bulgarian only imperfective verbs (complete-processes, terminative-actions) has specific «Present Participle» *използвам- използвАЙКИ , пиша- пишЕЙКИ* , which corresponds to the English ing-form: *use-usING, write- writING.*

2.5.3.2 Czech

Generally, in Agile, we do not try to derive the aspect counterparts. We allow multiple realisations of A-box concepts and the grammar then constrains the choice by lexical features perfective and imperfective. The entries in the file cadcam-dm-lex.lisp then will look as follows:

```
(annotate-concept DRAW :lex-items (KRESLIT NAKRESLIT))
```

Thus, for example, in the current version of the Czech lexico-grammar we only use the division to perfective and imperfective, but it may be changed in future, if we realise that we need a finer division.

2.5.3.3 Russian

```
ASPECT-TYPE
(clause-simplex) →
   [perfective] (process::perfective)
   [imperfective] (process::imperfective)
```

Note that in this case we do not use derivations for generation of the correct form in the lexicogrammar, hence the satement is not INFLECTIFY, rather CLASSIFY, which means presence of this feature at the lexical item which realises this process.

2.6 Clause Complexity

2.6.1 Introduction

The aim of the current chapter is to discuss the notion of clause complexity, and its possible comprehension in the context of Agile. To that end, we commence by briefly recapitulating Hallidayís views on clause complexity (Halliday,1994). We shall see that we can lift the perspective of the group as having a head and a modifier also to the level of clauses. Thus, similar to groups, we shall interpret sentences as clause complexes in which there is a head clause and other clauses that modify it. The idea of complexity then arises from the different ways (or dimensions) in which head clauses and modifying clauses can be related.

After the presentation of clause complexity in SFG, we continue by elucidating how the Nigel grammar accomplishes SFGís notion of clause complexity within the KPML framework. In particular, we shall have a look at NIGEL'S **clausecomplexity** region.

Given that we are concerned with Slavonic languages in the context of AGILE, the issue is in how far the approach taken in the NIGEL grammar can be adopted for generating sentences in Czech, Bulgarian, or Russian. In the final part of the chapter we make an assessment on the basis of corpus evidence, and show that in essence the setup of NIGEL can be used for Slavonic languages as well (as far as clause complexity is concerned).

2.6.2 Clause Complexity in SFG

The principle idea behind clause complexity in SFG (cf. (Halliday,1994)) is that we can view a sentence as a complex of clauses, just like a group can be perceived of as a **word complex**. And, similarly to word complexes, a **head clause** can be distinguished from clauses that **modify** that head clause. The assumption that SFG goes on is that the notion of clausecomplexity enables one to cover the entire spectrum of a sentence's (possible) functional organizations. As a matter of fact, Halliday defines a sentence *as a clause complex*, and the clause complex is the only grammatical unit recognized above the clause (cf. p.216, *ibidem*).

The complexity in clausal organization arises from the different ways in which clauses can be related. SFG interprets relations between clauses in terms of the 'logical' component of the linguistic system. In this interpretation two dimensions are distinguished, to be able to provide more detail on how exactly clauses can «modify» a head clause. The first dimension concerns the system of interdependency, being that of **parataxis** and **hypotaxis**[8]. The other dimension is the system of logico-semantics, regarding **expansion** and **projection**. In the remainder of this section we will describe these dimensions in some more detail. Crucial thereby is to bear in mind that clause complexity arises out of the interplay of both dimensions, (which is something we will examine in more detail in the section on NIGEL).

2.6.2.1 Interdependency or Taxis between Clause Complexes

We begin by discussing the system of **interdependency**, or **taxis**. The idea behind taxis is to elucidate the relative status of the head clause and the modifying clause(s). As Halliday puts it succincty (p.218, ***ibidem***), **hypotaxis** is the relation between a dependent element and its dominant, on which the former is dependent, whereas **parataxis** is a relation between two elements in which neither is dependent on the other (i.e. they stand on equal footing). It is the contention of SFG that all «logical» relations in language are either hypotactic or paratactic.

[8] As a matter of fact, the tactic dimension is general to all complexes, not just to clause complexes - in the same spirit it applies to word, group, and phrase.

We can consider a pair of clauses as a **clause nexus** relating a **primary clause** and a **secondary clause**. The following table shows associates these terms to taxis:

	Primary clause	Secondary clause
Parataxis	1 (initiating)	2 (continued)
Hypotaxis	α (dominant)	β (dependent)

Figure 15 Clause nexus / tactic dimensions

In the next subsections we discuss hypotaxis and parataxis in some more detail.

2.6.2.1.1 Hypotaxis

To repeat, **hypotaxis** is the binding of elements (here: clauses) of unequal status, namely a **dominant** and a **dependent** subordinate to the dominant. Due to the unequal status of participants involved in a hypotactic relation, the relation is logically speaking *asymmetric* and *non-transitive*. For example, consider the hypotaxis «I think when I program»: it follows that when I program, I think, but it is not necessarily the case that when I think, I must be programming. Naturally, this basic pattern can be modified by the nature of the logico-semantic relationship involved (see below) - but mostly, semantic properties are aligned as good as possible.

The elements in a hypotactic structure are ordered in dependence, but this order is largely independent of the linearity in the sentence. Thus, one can have a dependent clause (a) following the dominent, (b) preceding the dominant, (c) enclosed in the dominant, or (d) enclosing the dominant (cf. *ibidem*, pp.222-223):

(a)	The file is not stored	until you Save it.		
	α	β		α^β
(b)	If you Save the file,	then it will be stored.		
	β	α		β^α
(c)	Store,	if you want,	the file.	
	α	β	α	α «β»
(d)	You might,	the manual says,	save the file.	
	β	α	β	β «α»

2.6.2.1.2 Parataxis

In **parataxis**, the primary and secondary element of a clause nexus are of equal status. Both the **initiating** (primary) and **continuing** (secondary) element could stand as an individually functional whole - contrary to the case of hypotaxis, where the dependent cannot be considered outside the dominant it relies on. We also see this relative autonomy expressed in the logical characteristics of the paratactic relation, being *symmetry* and *transitivity*. Consider for example a simple conjunction like «manuals and cds». The conjunction «manuals and cds» implies «cds and manuals», (and vice versa), hence the relation is

symmetric. Furthermore, «manuals and cds» together with «cds and a box» can be taken to imply «manuals and a box», hence the relation is also transitive.

Because the initiating and continuing clause in a nexus are not dependent on one another, there is no ordering other than the linearity of the sentence - in contrast to hypotaxis, as we saw above.

2.6.2.2 Logico-Semantic Relations between Clause Complexes

We already indicated above that we characterize the modificationary relation between clauses in a clause complex not only in terms of the tactic dimension, but also in terms of **logico-semantic relationships**. SFG maintains a classification of logico-semantic relations that categorizes them using a small number of abstract types, in order to be able to handle the wide variety of relationships that can be distinguished as to hold between a primary and secondary clause in a clause nexus.

The classification is based on two fundamental classes, being (1) **expansion** and (2) **projection**. The notion of **expansion** expresses the idea that a clause, in its linguistic function as process, exchange and message, can enter into construction with another clause which augments it. The two clauses together then form a clause complex. The notion of **projection**, on the other hand, signifies those constructions in which the second clause is projected through the primary clause. The latter *instates* it as a locution, or an idea.

Because the Agile corpus does not show any clause complexes that can be classified as projections, we restrict ourselves to expansion in the discussion below. In essence, a secondary clause can expand the primary clause by **elaborating** it, **extending** it, or **enhancing** it. The remaining subsections elaborate these subtypes.

2.6.2.2.1 Expansion - Elaboration

In **elaboration**, the secondary clause develops further the meaning of the primary clause by specifying or describing it in more detail. As such, the secondary clause does not introduce a new referent in the discourse but rather provides a more elaborate description of an already present referent. Thereby it is not essential that the secondary clause is logically related to the primary clause as a whole - it may elaborate just a part of it (i.e. on one or more constituents).

In combination with parataxis, elaboration gives rise to the following three types:
1. **Exposition** – «in other words»
2. **Examplification** – «for example»
3. **Clarification** – «to be precise»

In an **exposition**, the secondary clause presents the viewpoint expressed in the primary clause in an other way, it may restate it in a different wording, or just strengthen it:

> «Windows © is stuck; it doesn't want to respond.»
> «KPML is a wonderful environment; it is just great.»

A secondary clause **examplifying** a primary clause develops the thesis of the latter in more detail, often by means of examples.

> *«You should regularly store your drawing –e.g. by clicking the Save button.»*

Finally, in case of **clarification**, the secondary clause elucidates the point made in the primary clause, by means of an explanatory comment.

«I wasn't surprised that Windows got stuck - it was what I had expected.»

Naturally, elaboration can also occur in conjunction with hypotaxis. This combination gives rise to what is called the category of **non-defining relative clause** in SFG. The meaning of the secondary clause is non-restrictive, or descriptive, with respect to the thesis expressed in the primary clause.

Note that **defining relative clauses**, unlike the non-defining ones, realise **embedding**. Embedding is functionally different from taxis as defined in SFG: *«Whereas parataxis and hypotaxis are relations BETWEEN clauses or other ranking elements, embedding is not. Embeddin gis a mechanism whereby a clause or phrase comes to function as a constituent WITHIN the structure of a group, which itself is a constituent of a clause. Hence there is no direct relationship between an embedded clause and the clause within which it is embedded.»* (Halliday 1994, p.242).

2.6.2.2.2 Expansion - Extension

In **extension** the secondary clause *adds* something new to the meaning expressed in the primary clause - contrary to the case of elaboration, above. What is added may either be a simple **addition**, or a **variation.** These are two variaties of **extension** (cf. *ibidem*, p. 230):

Category		Meaning
Additive		
	«and», additive: positive	X and Y
	«nor», additive: negative	not X and not Y
	«but», adversative	X and conversely Y
Variation		
	«instead», replacive	not X but Y
	«except», subtractive	X but not all X
	«or», alternative	X or Y

Figure 16 Types of extension

Paratactic extension is also known as **coordination** between clauses, typically expressed by *and, or, but*. The first subtype concerns addition. In this case, one process is simply adjoined to another, without there being any further causal or temporal relationship between them: *The system doesn't provide any explanation why it got stuck, nor would it be very helpful if it would.*

The second subtype of paratactic extension involves variation. Here, the meaning expressed by the second clause is put forward as a partial or total replacement of the thesis expressed by the primary clause: *They tried to do a good job, nevertheless the result was still of a deplorable quality.*

Again, we can discern subtypes of extension, depending on whether the relation occurs in combination with parataxis or hypotaxis. In conjunction with hypotaxis, extension yields addition, replacement, and alternation - however, in this case the secondary clause is of course dependent, as follows from the nature of hypotaxis. The dependent clause can be finite or non-finite (cf. (Halliday,1994) pp231ff.).

2.6.2.2.3 Expansion - Enhancement

Finally, in **enhancement** one clause enhances the meaning of another by qualifying it by reference to time, place, manner, cause or condition (i.e. **circumstantial**). We list the principal types of enhancement in the table below (cf. (Halliday,1994), p.234).

Category		Meaning
Temporal		
	same time	A meanwhile B
	different time: later	A subsequently B
	different time: earlier	A previously B
Spatial		
	same place	C there D
Manner		
	means	N by means of M
	comparison	N is like M
Causal-conditional		
	cause: reason	because P so result Q
	cause: purpose	because intention Q so action P
	condition: purpose	if P then Q
	condition: negative	if not P then Q
	condition: concessive	if P then contrary to expectation Q

Figure 17 Types of enhancement

Like the other two types of expansion, enhancement can be combined with either parataxis or hypotaxis. Paratactic enhancement yields a kind of coordination in which a circumstantial feature is incorporated, for example, *I like to follow up one line at a time, otherwise there is a muddle* (condition-neg)

A similar observation holds for hypotactic enhancement, though there we would speak of «adverbial clauses» involving a circumstantial.

2.6.3 Clause Complexity in Nigel

The way clause complexity is achieved in Nigel mirrors the treatment in SFG to a large degree. The region that is at the heart of generating clause complexity is **CLAUSECOMPLEX**. In this section we briefly describe (a relevant part of) the network that makes up this region.

The region's topmost system is CLAUSECOMPLEXITY which has as outputs the grammatical features **clausecomplex** and **clausesimplex**. The feature **clausecomplex** serves then as input for the system TYPE-OF-INTERDEPENDENCE, which is formally specified as follows:

```
type-of-interdependence:
(clausecomplex) →
      [expansion], [projection].
```

In the light of our discussion of clause complexity in SFG the reader may find the naming of the system a bit peculiar, given that SFG reserves the notion of interdependence for hypotaxis and parataxis. Because the AGILE corpus does not show any occurrence of projection, we do not consider that branch of the network any further.

Through **expansion**, two systems are triggered, namely EXPANSION-TYPE and EXPANSION-TAXIS. At various points in our discussion of clause complexity we eluded to the fact that SFG describes the complexity of modification in terms of an *interplay* of taxis and logico-semantic relations. The mentioned systems achieve exactly that:

```
Expansion-type:
    (expansion) →
       [enhancement]      (+Enhancement),
       [extension]        (+Extension),
       [elaboration]      (+Elaboration).

Expansion-Taxis
    (expansion) →
       [general-hypotactic-expansion],
       [paratactic-expansion].
```

The tactic dimension has its reflection in the graph structure through the following systems:

```
Parataxis:
    {paratactic-expansion,quoting} →
       [parataxis]
             (+Initiating,
              +Continuing,
              Initiating … Continuing).

Hypotaxis-Alpha-Complexity:
    (general-hypotactic-expansion) →
       [complex-alpha-hypotactic-expansion]
             (+Dependent,
              +Terminant,
              Terminant : full),
       [hypotactic-expansion]
```

On the first traversal through these systems, the basic type of expansion relation between the clauses in the nexus is determined, and the type of their interdependence. For the selected expansion relation a corresponding grammatical function is inserted in the graph structure, as is clear from the definition of the EXPANSION-TYPE system above, whereas the interdependency systems insert the elements relevant for the selected type of taxis (e.g. *initiating* and *continuing* for parataxis).

On a subsequent traversal, these «types» (or better, grammatical features) are combined to act as inputs to systems detailing out the exact kind of clause complexity. For example, **enhancement** together with **paratactic-expansion** form the input conditions for the system QUALIFYING-COORDINATION, whereas **enhancement** together with **general-hypotactic-expansion** serves as input conditions for the system QUALIFYING-CONDITION:

```
Qualifying-coordination:
    (enhancement, paratactic-expansion) →
       [qualifying-coordination]
             (+Enhanced,
              +Coordinator,
              Coordinator^Continuing).
```

1. *Tento rozměr definuje počátek 0,0 multičáry, relativně ke všem ostatním nakresleným elementům.*
   This measure defines the origin 0,0 of the multiline relative to all other drawn elements.

```
Qualifying-Condition:
    (enhancement, general-hypotactic-expansion) →
       [qualifying-condition]
             (Dependent : dependent-clause,
              Enhancement/Dependent).

Extending-Coordination:
    (extension, general-hypotactic-expansion) →
       [extending-coordination]
             (+Extended,
              Initiating/Extended,
              Continuing/Extension).

Non-defining-relative:
    (elaboration, general-hypotactic-expansion) →
       [nondefining-relative]
             (Elaboration : dependent-clause,
              Elaboration/Dependent).
    ⇒ ...Elaboration-extent [elaboration]

Apposition:
    (elaboration, paratactic-expansion) →
       [apposition]
             (+Elaborated,
              +Appositionmarker,
              Initiating^Appositionmarker,
              Initiating/Elaborated,
              Continuing/Elaboration,
              Elaborated : independent-clause,
              Elaboration: independent-clause).
```

In most cases, a further traversal through the region determines then the exact kind of clause complexity, for enhancement for example in terms of circumstantials (QUALIFYING-CONDITION, below). These systems introduce grammatical features that then act as input conditions for systems in regions of lower rank, concerned for example with the complexity of nominal groups (**NOMINALGROUPCOMPLEXITY**) or circumstantials (**CIRCUMSTANCE**). What is important about this observation is that the **CLAUSECOMPLEX** region is thus primarily concerned with *classifying* clauses, which to some extent influences the grammatical categories of the clauses involved, but leaving the more detailed realisation to other regions (indicated in the specifications by means of «⇒ ...»). The following formal definitions of systems in the **CLAUSECOMPLEX** region reflect this idea. The attentive reader will note that the systems below implement the different (sub)types of enhancement, elaboration and expansion as discussed in the previous section on SFG.

Systems detailing enhancement:

```
Qualifying-Coordination-Type:
    (qualifying-coordination) →
       [causal-coordination]
              (Enhancement/Initiating,
               Enhanced/Continuing,
               Coordinator ! So),
       [temporal-coordination]
              (Coordinator ! Then,
               Enhanced/Initiating,
               Enchancement/Continuing).

Qualifying-Condition-Type:
    (qualifying-condition) →
       [concessive-condition]
              (*Enhancement : concession-dependent)
       [conditional-condition] →
              (*Enhancement : conditional-dependent)
       [purposive-condition]
              (*Enhancement : purpose-dependent)
       [manner-condition]
              (*Enhancement : manner-dependent)
       [spatial-condition]
              (*Enhancement : spatial-dependent)
       [causal-condition]
              (*Enhancement : causal-dependent)
       [temporal-condition]
              (*Enhancement : temporal-dependent)
```

Systems detailing extension:

```
Extending-Hypotaxis:
    (extension, hypotactic-expansion) →
       [extending-hypotaxis] (Dependent/Extension).

Extending-Coordination-Type:
    (extending-coordination) →
       [additive-coordination]
              (+Coordinator,
               Coordinator^Extension,
               Coordinator ! And),
       [alternative-coordination]
              (+Correlator,
               Correlator^Extended,
               Correlator ! Either,
               +Coordinator,
               Coordinator^Extension,
               Coordinator ! Or),
       [contrastive-coordination]
              (+Coordinator,
               Coordinator ! But,
               Coordinator^Extension).
```

Systems detailing out elaboration:

```
Apposition-Type:
    (apposition) →
       [restating-apposition]
              (Appositionmarker ! I.E.),
       [exemplifying-apposition]
              (Appositionarker ! Eg).
```

Finally, important in relation to the discussion of transitivity elsewhere in this deliverable is the following system, triggered by **hypotactic-extension** or **clause-simplex**[9]. As is clear from the formal specification, the grammatical feature **transitivity-unit** serves as input to a large number of systems in other regions.

```
Transitivity-Unit:
    {clause-simplex,reporting,hypotactic-expansion} →
       [transitivity-unit]
             (+Predicator,
              +Process,
              +Lexverb,
              Lexverb/Process,
              Predicator (Lexverb),
              +Voice,
              +Predicator (Voice) ).
    ⇒ ... Temporal-Location [Circumstantial],
       Temporal-Extent-Adjunct [circumstantial],
       Spatial-Location [circumstantial],
       Spatial-Extent-Adjunct [circumstantial],
       Role-Adjunct [circumstantial],
       Matter-Adjunct [circumstantial],
       Manner-Adjunct [circumstantial],
       Cause-Adjunct [circumstantial],
       Accompaniment-adjunct [circumstantial],
       Process-Type [nonrelational-transitivity],
       phase [nonrelational-transitivity].
```

2.6.4 Evidence from the Corpus

2.6.4.1 Introduction

In this section we present examples of how clause complexity manifests itself in the AGILE corpus. The way we go about is by providing for each of the clause complex types discerned in SFG (i.e. taxis × logico-semantic relations) sets of examples from the three Slavonic languages considered in AGILE.

On forehand we should mention that the Agile corpus (or, more specifically, the language-specific corpora) does not contain any manifestation of clause complexity involving **projection**. For this reason we do not consider it in the presentation below.

Each example is presented with its reference(s) to its location in the (language-specific) corpus. Furthermore, for each clause complex type we note the tags that determine it in the corpus (following the discussion in WP 3).

2.6.4.2 Corpus Data

2.6.4.2.1 Elaboration + Parataxis

2.6.4.2.2 Exposition

Czech, Bulgarian, Russian: No occurrences

[9] Or **reporting** - which, however, is irrelevant to our discussion here, since this grammatical feature falls under projection.

2.6.4.2.2.1 Exemplification

Czech:

- (65) Můžete zadat tyto relativní hodnoty ve formě @vzdálenost<úhel ((66)v tomto případě, raději zadejte @**3<100**).
  You can enter these relative values in the form @distance<angle (in this case, you would enter @3<100).
- (103) Například, jestliže vytvoříte element s negativním offsetem**,** (104) objeví se pod originálem v obrazovém políčku.
  For example, if you create an element with a negative offset, it appears under the original in the window.
- (117) Například, pokud máme šest elementů, (118) vnitřní oblouky spojí elementy 2 a 5 (119) a elementy 3 a 4.
  For example, if we have six elements, the inner arcs will connect elements 2 and 5, and elements 3 and 4.

Bulgarian:

- (49) Можете да въведете тези относителни стойности под формата на @разстояние<ъгъл (в дадения случай: @3<100).
  You can enter these relative values in the form @distance<angle (in this case, you would enter @3<100).

Russian:

- (58) Можно указать эти данные в следующей форме: @расстояние<угол (в нашем случае: @3<100).
  You can enter these relative values in the form @*distance<angle* (in this case, you would enter @3<100).

2.6.4.2.2.2 Clarification

Czech:

- (10) Pro ukončení úsečky (11) stiskněte ENTER, (12) nebo zadejte **u** (13) pro uzavření, (14) t.j. pro spojení počátečního bodu prvního segmentu s koncovým bodem (6) posledního segmentu**.**
  In order to end the line, press Enter or enter **u** for closing, i.e. for connecting the start point of the first element with the endpoint of the last element.

Bulgarian, Russian: No occurrences

2.6.4.2.3 Elaboration + Hypotaxis

All the three Slavonic languages allow for two ways of expressing hypotactic elaboration, one is by a relative clause, the other by a participal modifier. It should be explicitly noted here again that hypotactic elaboration involves **non-defining (non-restrictive) relative clauses or by non-restrictive participal modifiers**.

Relative subordinate clauses The structure of relative subordinate clauses in all the three Slavonic languages is similar to English; the relative pronoun agrees with its antecedent in

number and gender (case may of course be different, because it is governed by the verb in the relative clause as whose complementation the relative pronoun functions).

Participial modifying clauses can also be used instead of relative clauses in the Slavonic languages.

Czech:
Non-restrictive Relative clauses

- (41) ÚSEČKA vytváří jednoduché nebo vícenásobné úsečkové segmenty, (42) které jsou samostatné objekty.
  LINE creates a single or multiple line segments which are individual objects.

Non-restrictive Participle modifying clauses are not frequent in Czech:

- (48)Úhel vyplnění je úhel vytvořený pomocí třetího bodu, který určíte a druhého bodu a kladné části osy X.
  The angle of the filling is an angle created using a third point you specify and the second point and the positive part of the X axis.

Bulgarian: No occurences

Russian:
Non-restrictive Relative clauses

- Командой LINE создаются единичные или множественные линейные сегменты, [39] которые являются отдельными объектами [40].
  LINE creates single or multiple line segments (Acc, pluaral) that (Acc, pluaral) are separate objects.

Participial modifying clauses. This type of clauses is frequently used in Russian instead of a relative clause or a circumstantial prepositional phrase without a major process in English:

- Команда PLINE позволяет рисовать сегменты полилиний и дуг, образующие единый объект [22].
  PLINE draws poliline line and arc segments that form (In Russian - «forming») a single object.
- Команда OFFSET создает копии линий, смещенные на указанное расстояние в одну сторону от точки или через точку [24].
  OFFSET creates copies of lines offset at a specified distance (In Russian - «moved for a specified distance o») to one side or through a point.

Embedding

We also include some examples of defining (restrictive) modification by relative clauses and participles here, even though they are a case of embedding not hypotaxis. Embedding belongs to the nominal group complex. We exemplify them both here because the two constructions are the same syntactically and so might be easily confused in on eis not careful.

Czech:

Restrictive Relative clauses

- Zadejte kdykoli **z** a čára, kterou právě kreslíte, a ty, které již byly nakresleny. budou zapsány do databáze
  Enter **z** any time and the line which you have been drawing, and those that you have already drawn, will be entered into the database.

Restrictive Participle Modifiers

- (22) KŘIVKA kreslí polyúsečku složenou ze segmentů úseček a oblouků, (23) jejíž forma je jeden objekt.
  PLINE draws a polyline consisting of line and arc segments whose form is one object.
- (234) Jestliže je pero dole, (235) můžete pokračovat v kreslení po provedeném záznamu.
  If the pen is down, you can continue drawing after recording is done.

Bulgarian:

Restrictive Relative clauses

- (192) за да запазите в базата данни линията, (и линиите,) (193) която чертаете (194) които вече сте начертали.
  to record (save) in the database the line you've drawing and those already drawn.
- (218) Преместете показалеца в края на линията, (219) която сте скицирали последна,
  Move the cursor to the end of the line you drew last

Restrictive Participle Modifiers

- (171) Скициране и записване на линии, начертани на ръка
  To sketch and record freehand lines

Russian:

Restrictive Relative clauses

- *... до конца фрагмента (Gen, sing, masc), [203] который (Gen, sing, masc) вы хотите стереть [204].*
  to the end of a fragment (Gen, sing, masc) that (Gen, sing, masc) you want to erase.
-

Restrictive Participle Modifiers

- (191) *линию, рисуемую*(Acc, sing, femin) *в данный момент*
  the line you are drawing

2.6.4.2.4 Extension + Parataxis

*2.6.4.2.4.1 Addition (**PRC,CTAD**)*

***Czech**:*

- *(110) Pod Zakončení zvolte úsečku nebo oblouk pro každou multičáru (111) a zadejte úhel.*
  Under *Finishing* choose line or arc for each multiline and specify angle.
- *(224) Stiskněte znova výběrové tlačítko (225) a pero se zdvihne tak, (226) že nyní můžete pohybovat kurzorem po obrazovce (227) bez kreslení.*
  Press the choice button again and the pen will lift so that now you can move the cursor on the screen without drawing.

Bulgarian:

- *(84) В подменю Caps изберете линия или дъга за всеки край на мултилинията (85) и въведете ъгъл.*
  Under Caps, select a line or an arc for each end of the multiline,
  and enter an angle.

Russian:

- *[88]В разделе Caps укажите прямую или дугу для каждого конца мультилинии [89] и введите угол.*
  Under Caps, select a line or an arc for each end of the multiline, and enter an angle.

*2.6.4.2.4.2 Variation (**PRC,CTAL**)*

Czech:

- (69) Poté co vytvoříte křivku, (70) ji můžete editovat pomocí příkazu KEDIT (71) nebo použít příkaz ROZLOŽ (72) pro přeměnění křivky na jednotlivé úsečky a obloukové segmenty.
  After you have created a polyline, you can edit it using the KEDIT command or use the UNGROUP command in order to convert the polyline into individual lines and arc segments.
- (173) Určete čtvrtý bod (174) nebo zadejte **U** (175) pro uzavření multičáry (176) nebo stiskněte ENTER (177) pro její ukončení.
  Specify the fourth point or enter **u** to close the polyline or press ENTER to end it.

Bulgarian:

- можете да я редактирате с PEDIT или да използвате EXPLODE
  you can edit it with PEDIT or use EXPLODE …
- (86) Линиите пресичат края на цялата мултилиния,
  (87) а външните дъги съединяват краищата на най-външните елементи.
  Lines cross the end of the whole multiline and outer arcs join the ends of the outermost elements.

- (214) Въведете **е** от Erase независимо от това, (215) дали „моливът« е вдигнат или поставен.
  …with the pen up or down

2.6.4.2.5 Extension + Hypotaxis

2.6.4.2.5.1 Addition

Czech, Bulgarian, Russian: No occurrences

2.6.4.2.5.2 Variation

Czech, Bulgarian, Russian: No occurrences

2.6.4.2.6 Enhancement + Parataxis

*2.6.4.2.6.1 Temporal (**PRC, CTT**)*

Czech:

- (44) V dalším příkladu si nakreslíte rovný segment křivky, (45) na který navazuje obloukový segment, (46) pak nakreslíte jiný rovný úsek v tečném směru.
  In the next example you will draw a straight line segment, which is followed by an arc segment and then you will draw another straight segment in the perpendicular direction.

Bulgarian:

- (219) Преместете показалеца в края на линията, (220) която сте скицирали последна, (221) и след това го върнете по линията до мястото
  Move the cursor to the end of the line you drew last and then move it back…

2.6.4.2.6.2 Causal-Reason

Czech:

- (218) Stiskněte výběrové tlačítko (219) a spusťte pero dolů.
  Press the choice button and lower the pen.

Bulgarian, Russian: No occurrences

2.6.4.2.6.3 Causal-Condition

Czech

- (224) Stiskněte znova výběrové tlačítko (225) a pero se zdvihne tak, (226) že nyní můžete pohybovat kurzorem po obrazovce (227) bez kreslení.
  Press the choice button again and the pen will lift so that you can now move the cursor on the screen without drawing,

2.6.4.2.7 Enhancement + Hypotaxis

*2.6.4.2.7.1 Temporal (**HRC, CTT**)*

Czech:

- (206) Poté co jste vytvořili polygon, (207) můžete jej editovat pomocí příkazu KEDIT (208) nebo jej zkonvertovat na jednotlivé úsečkové segmenty (209) příkazem ROZLOŽ.
  After you have created a polygon, you can edit it using the KEDIT command or convert it into individual lines and arc segments using the UNGROUP command.
- (239) Když stisknete tlačítko, (240) kreslení od ruky začne, (241) ať je kursor kdekoli.
  If you press a button, sketching will start no matter where the cursor is.

Bulgarian:

- (168) След като сте създали многоъгълник, (169) можете да го редактирате с PEDIT
  After you've created a polygon, you can edit it with PEDIT
- (227) След като вече сте записали линиите, начертани на ръка, (228) не можете да ги редактирате
  Once you record freehand lines, you can't edit them

Russian:

- После того, как шестиугольник нарисован [172], его можно редактировать с помощью команды PEDIT [173]
  After you've created a polygon, you can edit it with PEDIT
- Нажимайте клавишу «r» каждый раз [190], когда хотите записать (сохранить) в базе данных линию [191],
  Enter r at any time to record (save) in the database the line

«Adverbial-participial clauses» (called «přechodník« in Czech and «деепричастный оборот» in Russian)

The standard mean to express simultaneity in Russian is the adverbial-participial clause (it is rankshifted to the Circumstance level; its subject being the same as the subject of the main clause). This corresponds closely to the English gerund constructions.

- Записывая произвольную линию, вы не можете ее редактировать или стирать опцией Erase.
  While recording freehand lines, you can't edit them or erase them with the Erase option.

*2.6.4.2.7.2 Spatial (**HRC,CTS**)*

Czech:

- (73) Při rozkládání široké křivky (74) se šířka čáry změní na 0 (75) a výsledná křivka je umístěna podél středu (76) kde byla křivka.
  When decomposing a thick line, the width of the line changes to 0 and the resulting line is placed alongside the centre where the line was.

Bulgarian:

- (201) Линията започва от мястото, (202) където се намира показалецът в момента на натискане на бутона.
  The freehand line starts wherever the cursor is

*2.6.4.2.7.3 Manner (**HRC,CTM**)*

Czech:

- (17) Novou úsečku můžete začít v koncovém bodu (6) poslední nakreslené úsečky (18) stisknutím ENTER
  You can start a new line in the end point of the last drawn line by pressing ENTER.

It is common in Czech to express manner by the following construction, which however does not appear in the corpus:

- Spusťte příkaz ÚSEČKA použitím jednoho z těchto způsobů*:*

There are the following cases which we now classify as clause complexity, but they might belong to group complexity rather than clause complexity.

- (87) Tento rozměr definuje počátek 0,0 multičáry, relativně ke všem ostatním nakresleným elementům.
  This measure defines the origin 0,0 of the multiline relative to all other drawn elements.
- (224) Stiskněte znova výběrové tlačítko (225) a pero se zdvihne tak, (226) že nyní můžete pohybovat kurzorem po obrazovce (227) bez kreslení.
  Press the choice button again and the pen will lift so that you can now move the cursor on the screen without drawing,

Bulgarian:

- (2) Стартирайте командата LINE, (3) като използвате един от следните методи:
  Start the LINE command , using one of the following methods
- (88) Вътрешните дъги свързват двойки от елементи, (89) като оставят средната линия несвързана
  Inner arcs connect pairs of elements, leaving the centerline unconnected ...

Russian:

Adverbial-participial clauses in Russian typically express manner:

- Запустите команду MLINE, выбрав Multiline в палитре Polyline на панели инструментов Draw.
  Start the MLINE command by choosing Multiline from the Polyline flyout on the Draw toolbar.

*2.6.4.2.7.4 Clausal-Condition (**HRC, CTCC**):*

*2.6.4.2.7.5 **Czech**:*

- (103) Například, jestliže vytvoříte element s negativním offsetem, (104) objeví se pod originálem v obrazovém políčku.
  For example, if you create an element with a negative offset it will appear under the original in the window.
- (234) Jestliže je pero dole, (235) můžete pokračovat v kreslení po provedeném záznamu.
  If the pen is down, you can continue drawing after recording is done.

Bulgarian:

- (89) като оставят средната линия несвързана, (90) ако броят на елементите е нечетен.
  leaving the centerline unconnected if there is an odd number of elements.
- (216) Ако „моливът« е поставен, (217) той се вдига сам.
  If the pen was down, it moves up.

Russian:

Russian has in this case a special taxis of tenses in main and subordinate clauses; it is related to the meaning of achievements referred to by the perfective aspect, so a form in the past tense, passive voice is used in the subordinate clause and a form in the future tense (morphologically the present tense of the perfective aspect), active voice in the main clause:

- Если перо было опущено, [200] оно автоматически поднимется [201].
  If the pen was down, it moves up.

*2.6.4.2.7.6 Clausal-Purpose (**HRC, CTCP**):*

Czech:

- (259) Abyste kreslili čáry od ruky jako křivku, (260) která má být samostatný objekt, (261) nastavte systémovou proměnnou SKPOLY (262) před kreslením na nenulovou hodnotu.
  In order to sketch a line as a POLYLINE, which should be an individual object, set the system variable SKPOLY to 0 before drawing.
- (10) Pro ukončení úsečky (11) stiskněte ENTER, (12) nebo zadejte **u** (13) pro uzavření, (14) t.j. pro spojení počátečního bodu prvního segmentu s koncovým bodem (6) posledního segmentu.
  In order to end the line, press Enter or enter **u** for closing, i.e. for connecting the start point of the first element with the endpoint of the last element.

Bulgarian:

The most frequent case:

- (11) Натиснете Return, (12) за да завършите линията
  Press Return to complete the line

- (91) Ако искате да видите основния цвят, (92) изберете On в подменю Fill
  Under Fill, select On to display a background color

Russian:

The most frequent type of subordinate clauses in the corpus (38 occurrences) is the infinite purposive clause expressing the purpose of a sequence of operations by means of the special connector *чтобы* and the infinitive rankshifted clause:

- Чтобы нарисовать линию *[1]*
  To draw a line (p. 44)
- Запустите команду LINE [2] одним из следующих способов:[3]
  Start the LINE command using one of these methods:

or the purpose of a simple procedure:

- 5 Нажмите клавишу Return, [11] чтобы завершить рисование линии, [12]
  5 Press Return to complete the line, ... or

In the latter case, it is used inside of a clause complex. The order of clauses can be direct (see above) or inverse:

- Чтобы отменить последний линейный сегмент в команде LINE,[16] нажмите клавишу «u» *[17].*
  To undo the previous line segment during the LINE command, enter u.

This follows the English order, so it is related to communicative structure.

Purpose can be expressed also by nominalization of the process and using it in a circumstantial prepositional phrase «для + Noun (Genitive)»:

- 5 Нажмите клавишу «c» [13] для запуска команды Close [14],
  5 Enter c for Close

In contrast to English for Close, the process *запуск* (start) is explicitly mentioned: *for starting the Close command.*

Where the process cannot be expressed by a nominal group, a heavy subordinate clause is necessary:

- 4 Введите «i» [153] для указания на то, что многоугольник вписан в круг [154].
  4 Enter i for Inscribed in circle. (Enter i for specification that the polygon is inscribed in the circle).

2.6.4.2.8 Some Connectives

The table below illustrates the connectives encountered in the corpora:

Clause complex types	Bulgarian	Czech	Russian	English
Enhancing + Hypotaxis: Clausal-purpose	за, за да	aby	Чтобы	to, for
Enhancing + Hypotaxis: Temporal overlap	Когато	když,	после того, когда	after, when
Enhancing + Hypotaxis: Temporal sequence	След като, след като вече	až potom, pak	∅	once, by ...
Enhancing + Hypotaxis: Clausal-conditional	ако	jestliže, pokud, když	Если	if
Elaboration + Hypotaxis: (Non-restricitve Relative subordinate clauses)	който, която, което, които	který, jenž, jehož	Который	that, which, ∅
Extension-Variation (alternative) + Parataxis	или	nebo	Или	or
Parataxis in general	и, а, но	a	И	and

Figure 18 - Connectives

2.6.5 Conclusions

We carried out a detailed study of the various types of clause complexity manifested in our corpora, elaborating on our corpus analysis presented in the deliverable of WP3 (corpus analysis). Given the types of relations between clauses that are distinguished in SFG and in the NIGEL grammar, it has turned out that:

1. A wide range of types is manifested in the corpora of the three languages – in fact, practically all types of *expansion* are present.
2. The SFG-based typology is sufficient to cover the relations we find in our languages, which is in itself not surprising given that they are meanings and not realisations.

Differences are encountered between English and the Slavonic languages as well as between Czech, Russian and Bulgarian, in some of the realisations of some of the particular relation types.

In general, there is correspondence between English and the Slavonic languages in expressing paratactic relations, irrespective of the underlying locical relation. The language-specific equivalents of the connective 'and' serve this general purpose.

When hypotactic relations are considered, there is also a range of constructions that correspond across languages. We did not encounter any occurrences of hypotactic extensions, so we leave them out in this discussion.

As for hypotactic elaboration, the ways to express it are the same in the Slavonic languages as in English, namely by non-restrictive relative clauses. Unlike in English, the relative pronoun exhibits agreement with its antecedent in gender and number, its case is

governed by the verb in the relative clause. The Slavonic languages, and especially Russian, use also non-restrictive participle modifiers for hypotactic elaboration. In Czech written text it needs to be separated by a comma to distinguish it from a restrictive modifier. In Russian subordinate clause (finite form) is always separated by commas at the beginning and the end.

Regarding hypotactic enhancement, we can say that there are no big differences in expressing spatial relations. In expressing temporal relations, we encounter a difference between Czech and Bulgarian on the one hand, and Russian and English on the other hand. English can use gerunds for temporal overlap or sequence (*While recording* | *Having chosen* Multiline *you can* …). Russian can use the present or past participle clause for this purpose, with restrictions on the combination of tense and aspect in the main and the subordinate clause (*Записывая произвольную линию, вы можете…*| *выбрав Multiline вы можете* …). Such constructions also exist in Czech, but they have become uncommon in the modern language. Technical texts still do contain it and it is not considered an anachronism when such way of expression enables to condens the text.

Also hypotactic clausal condition is very similar in the Slavonic languages mututaly and with respect to English. In Russian we have to take into account the tense and aspect interplay.

The most variation across languages is found in expressing Clausal Manner and Clausal Purpose. Clausal Manner is expressed either by a finite subordinate clause in Czech, Russian and Bulgarian. In Czech and Russian, also nominalization corresponing to the English *by –ing* construction is available, though no preposition is used, just the instrumental case. In Czech, also a construction corresponding to the English *using* is available. In Russian it is common to use the participle clause to express manner.

Likewise, there are interesting variations in expressing Clausal Purpose. In all three languages, it is possible to use a subordinate purpose clause. In Czech it is a conditional clause, in Russian an infinitival clause, in Bulgarian a finite subordinate clause. These all are used in place of the English *in order to*. Besides that, Czech and Russian use also a nominalization in a prepositional phrase, which can be glossed in English as *for –ing*. In Czech, this construction *pro+nominal* is perceived as marginally grammatical by native speakers, however, it is becoming more and more frequent together with other uses of prepositional phrases with *pro*, probably due to the influence of languages like English and German which use the *for* (*für*) much more frequently.

The various ways of realizing the individual types of clause complexes depending on the logical relations underlying them is one of the foci of grammar development in the current stage of the AGILE project. The decision for the complexity of a sentence is to be made at the level of *text structuring*. The decision mechanisms to be employed thereby will be addressed and reported in the context of the WP5 work package.

2.7 Word Order

Communication in natural language is a process in time: a speaker (writer) can only produce and a hearer (reader) can only receive the elements one by one. Linearity is a universal characteristic of the natural language communication process. Therefore, the content that is being communicated, which is by nature multidimensional, needs to be conveyed in a **linear order**. Linguistic structures of all degrees of complexity are projected into one dimension, and the individual elements are ordered sequentially according to certain rules. The rules and schemes of linear ordering can be studied at every stratum of a given language, including the

lowest ones, i.e. graphemes, phonemes, morphemes etc. The rules and schemes pertaining to the linear ordering of elements constituting a clause, i.e. phrases and groups, is usually referred to as **word order** (even though the elements are not only individual words). The linear ordering of clauses within a complex clause (sentence) can be referred to as **clause order**.

Various linguists have discerned certain factors in the language system in general which, besides their other functions, play an important role in expressing a given content in a linear form. The inventory of these factors contains at least the following: **information structure, grammatical structure, intonation, rhythm and style**. These factors are very general, and can therefore be considered language universals, at least within the family of indo-european languages. These languages more or less share the inventory of the factors. However, they may differ significantly in the hierarchy of these factors: the individual factors may have different importance for the linear ordering, i.e. word order and clause order, in a given language. The hierarchy is determined by the systems of individual languages.

Information structure is considered the main factor determining the linear ordering within a sentence in the Slavonic linguistic tradition. It is considered relevant for both word order and clause order. Hereby, we are using the term information structure as a general term for various notions employed in contemporary theories of the syntax-semantics interface, notions that reflect how the conveyed content is distributed over a sentence, and how it is thereby structured or «perspectivized».

Within the Czech linguistic tradition, information structure has been referred to as «**aktuální členění**» (Sgall et al. 1980), **functional sentence perspective** (Sgall 1967, Firbas 1992) and **topic-focus articulation** (Sgall et al. 1973, 1986). Another terminology was introduced by Chafe in the 70's and used in (Vallduví 1992), where information structure is called **information packaging**.

Halliday (1967; 1970) makes a distinction between the **thematic structure** of the sentence and its **information structure**; he distinguishes between 'given' and 'theme', where given corresponds to «a point of contact with what the hearer knows», whereas 'theme' corresponds to «the heading to what the speaker is saying».

Even though different theories use different terminology, it has by now been generally established that information structure is relevant semantically, not only pragmatically.

In every approach to information structure, the clause is considered to consist of (at least) two parts. The often used oppositions are, e.g. Theme-Rheme (Sgall 1973, Halliday 1985, Daneš, 1985, Steedman 1991), Topic-Focus (Sgall et. al. 1986, Partee et al. 1998), Background-Focus (Krifka 1992), Ground-Focus (Vallduví 1992). Sometimes, the authors introduce further sub-divisions, e.g. Firbas distinguishes Transition as a part of the Rheme, Vallduví considers the Ground to consist of a Link and Tail. For a discussion of some of the differences between these approaches see (Hajičová and Korbayová, 1997).

In the current project, we cannot consider all these different approaches to information structure. In the end, we have to work with just one approach for the sake of our linguistic specifications. Henceforth, we restrict ourselves to two candidate approaches: Halliday's thematic structure as developed in the Systemic Functional Grammar framework (SFG) and the topic-focus articulation approach developed in Prague within the framework of Functional Generative Description (FGD). Halliday's SFG approach is chosen because the current treatment of information structure in the KPML system is based on it. The FGD

approach serves us to elaborate the SFG approach towards a more flexible treatment required for languages with a higher degree of free word order than English.

The main questions to be addressed in this part of the present report are as follows:

- how to control word order in the course of generation
- what are the factors that determine word order in Czech, Russian and Bulgarian
- how can we capture the word order regularities in the SFG framework
- and subsequently, how can we reflect them in KMPL

The discussion is organized as follows. We first present some examples of word order variations in simple sentence in the three languages (Section 2.7.1). These examples illustrate that differences in word order very often correspond to differences in **information status** of the entities and processes about which the text is, in particular whether they are already familiar or not, and whether they are assumed to be salient for the addressee. But also, word order can reflect text organization which have to do with the way how some information is to be presented to the addressee.

Considerations of this kind have been reflected in the SFG framework In particular, SFG recognizes the **thematic structure** and the **information structure**, and they both have to do with ordering of constituents with respect to the larger context. We overview the SFG terminology and discuss an SFG-based approach to word order in Section 2.7.2. Using solely the apparatus available in SFG is, however, not sufficient to cover the facts concerning word order in our three languages.

In order to achieve a more complete account, we use some of the ideas developed within the framework of Functional Generative Description in Prague. We explain the essential notions and try to incorporate them into the SFG-based approach in Section 2.7.3.

We propose an ordering algorithm based on the combined insights (Section2.7.4). We preserve the SFG notion of Theme as being the first experiential element in the clause and being determined by text organization, because it appears useful also in our three languages. For ordering of non-thematic constituents within a clause we use notions adopted from FGD, namely the so-called systemic ordering in combination with the distinction between contextually bound and non-bound elements. Contextually bound elements are ordered separately from the non-bound ones. The ordering of the latter follows systemic ordering. The ordering of the former does not have to follow systemic ordering, but we use systemic ordering as a default. The ordering of clitics with respect to other elements is special, namely they occupy the so-called Wackernagel's position, i.e. second. In the present treatment, the second position is the position immediately following the Theme.

We conclude with a discussion of the overall coverage of the proposed approach with respect to the realisation of information structure by word order (Section 2.7.5). We tackle the question whether it is sufficient to consider textual phenomena as «second order» ones with respect to the ideational and interpersonal phenomena, i.e. whether it suffices to see information structure and thus also word order as operating upon grammatical structure generated prior to taking the textual perspective into account.

2.7.1 Description of phenomena

In this section, we discuss some examples of word order variation in Czech, Russian and Bulgarian, and try to show the differences in meaning that are reflected in this way. We make an initial comparison across the three languages,

Slavonic languages are generally considered to exhibit a **high degree of free word order**. This characteristic is based on their comparison with languages like English or French, where clause constituents cannot be «moved around» with the same relative freedom without simultaneous changes in the syntactic structure. To illustrate this phenomenon, let us consider the Czech sentence in (2) and the permutations of its main syntactic groups in the following example (sentence (1) provides the context, (1) and (2) are extracted from our intermediate demonstrator text)[10]:

Czech

(1) *Otevřete dialogový panel* Styly multičár *jednou z následujících metod.*
Open dialog pane-acc Styles Multiline *one*-instr *of following methods*-gen
Open the Multiline styles *dialog box using one of the following methods*

(2) *Z menu Data vyberte* Styl Multičáry.
From menu-gen Data *choose* Style-acc Multiline-gen
From the Data *menu choose* Multiline Style.

(a) *Vyberte z menu* Data Styl Multičáry.
choose from menu-gen Data Style-acc Multiline

(b) *Vyberte Styl Multičáry z menu Data.*
choose Style-acc Multiline from menu-gen Data

(c) *Styl Multičáry vyberte z menu Data.*
Style-acc Multiline choose from menu-gen Data

(d) *Z menu Data Styl Multičáry vyberte.*
From menu-gen Data Style-acc Multiline choose

(e) *Styl Multičáry z menu Data vyberte.*
Style-acc Multiline from menu-gen Data choose

All the permutations of the main syntactic groups in sentence (2) shown in (2a) – (2e) constitute grammatically well formed sentences of Czech. It is therefore apparent that there is a big variability of word order in Czech, that is, for sentences with a moderate number of syntactic groups there are usually quite a few grammatically well formed word order variants. The possibility to form many word order variants is given by the language system of Czech and is one of the characteristics of the Czech word order. However, one should not interpret the high degree of freedom in word order as arbitrariness. The word order of a Czech sentence is free, but not arbitrary.

While some word order variants can be interchanged in a context, this is certainly not true of all possible variants. With respect to the example above, not all of the variants in (2a) – (2e) would be acceptable in the context of our intermediate demonstrator text, i.e. in the context of sentence (1): it appears that both versions (2a) and (2b) could be used instead (2); however, the remaining version (2c) – (2e) could not. In fact, (2d) and (2e) could only be used felicitously in a rather restricted type of contexts, namely those where the verb or the

10 In all the examples discussed in the present report, we assume an «unmarked» intonation pattern, i.e. intonation center at the end of a clause (or other prosodic unit).

action referred to is to be interpreted as contrasting some other verb or action with the same participants and circumstances. (2c) does not fit into the context of (1) simply because it presupposes **salience, i.e. contextual familiarity**, of «Styl multičáry», which is not granted in this context. Also, let us remark that the version in (2a) is perceived as **neutral** in Czech, or unmarked in a general sense, out of context. It does not make any specific requirements on the context in which it can be used felicitously.

In Russian, not all the permutations constitute acceptable sentences.

Russian

(3) *Откройте диалоговое окно Multiline Styles*
Open dialog window-acc Multiline Styles
одним из следующих способов
one-instr from following methods-gen
Open the Multiline styles dialog box using one of the following methods

(4) *В меню* Data *выберите пункт* Multiline Style.
In menu-gen Data *choose item* Multiline Style
From the Data menu choose Multiline Style.

(a) *Выберите в меню Data пункт* Multiline Style.

(b) *Выберите пункт Multiline Style в меню Data.*

(c) °*Пункт Multiline Style выберите в меню Data.*

(d) °*В меню Data пункт Multiline Style выберите.*

(e) °*Пункт Multiline Style в меню Data выберите.*

Similarly to Czech, the variants (4a) and (4b) could be used instead of (4) in the context of (3), but (4c) is not appropriate. The choice between (4), (4a) and (4b) depends on the text style and contextual conditions. The default order in imperatives in Russian is Verb<Patient<Locative.

A sentence like (4c) is not encountered in the AGILE corpus, though it may be used to express an emphasis on the location of the item in a contrastive context. (4d) and (4e) sound really impossible to be used for any purpose, though still being grammatical.

It holds for both Russian and Czech that there are sentences where it would be entirely natural and appropriate to put the verb into the final position, for example:

(5) *А теперь узлы завяжите (Russian)*
A nyní uzly zavažte. (Czech)
And now knots-acc tie-up
And now tie the knots up.

(6) *А теперь файл сохраните (Russian)*
A nyní soubor uložte. (Czech)
And now file-acc save
And now save the file.

The general explanation can be that a sentences such as (5) and (6) can be used in a context where the knots or the file have already been talked about and the sentence expresses the next action to be performed with them, i.e. to tie them up or to save it. But in Russian this presupposes a heavy salience of the verb. Another interpretation is with

intonational salience of *файл*. It will be interpreted as if there was something already saved and the *файл* is the next thing to be saved . But intonation of the sentence in Russian is not neutral anyway.

The translation into Bulgarian in shown in (7) and (8) below. As in Czech and Russian the variants in (8), (8a) and (8b) are acceptable in the context of sentence (7), while (8c) is out-of-context. The word order in (8e) and (8d) results in ungrammatical sentences (such word order may be met potentially only as inversion in poetry).

(7) *Отворете диалоговия прозорец Multiline styles*
Open dialog window Multiline styles
с един от следните методи.
with one of following methods.
Open the Multiline styles dialog box using one of the following methods

(8) *От менюто Data изберете Multiline Style*
From menu-def Data choose Multiline Style.
From the Data menu choose Multiline Style.

(a) *Изберете от менюто* Data Multiline Style.

(b) *Изберете Multiline Style от менюто Data.*

(c) *Multiline Style изберете от менюто Data.*

(d) * *От менюто Data Multiline Style изберете.*

(e) * *Multiline Style от менюто Data изберете.*

The above examples in Czech, Russian and Bulgarian illustrate that sentences which differ only in word order (and not in syntactic realisations of constituents) are not freely interchangable in a given context. The comparison of the possible permutations for each language also illustrates that the degree of word order freedom is not the same in across the three languages under consideration. In order to explain the differences in meaning that are often reflect by differences in word order, let us now consider another example in all three languages:

(9) *Czech*

(a) *Otevřete soubor příkazem Open.*
Open-imp file-acc command-instr Open
Open a|the file by the Open command.

(b) *Soubor otevřete příkazem Open.*
File-acc open-imp command-instr Open.
Open the file by the Open command.

(c) *Příkazem Open otevřete soubor.*
command-instr Open open-imp file-acc
By the Open command open a file.

(d) *Příkazem Open soubor otevřete.*
command-instr Open file-acc open-imp
By the Open command open the file.

(e) *Otevřete příkazem Open soubor.*
open-imp command-instr Open-imp file-acc
Open a file by the Open command.

(f) *Soubor příkazem Open otevřete.*
File-acc command-instr Open open-imp.
Open the file by the Open command.

The sentence in (9a) is neutral. It can be used «out of the blue», or in a context which can be approximated by the question *What should we do?* There are no implicatures concerning the presence of a file, or the identity of the file to be opened.

The sentence in (9b) is appropriate when some file is salient, for instance when the user is working with a file, because it means opening of a specific file. That is why we put the definite article into the English translation. The action of opening can, but does not have to, be salient, too. The contexts in which (9b) can be appropriately used can be characterized by the questions *What should we do with the file?* or *How should we open the file?*

Also (9f) presumes salience of a file, and also of the Open command. The contexts in which (9f) can be appropriately used can be characterized by the questions *What should we do to the file with the* Open *command?*

The verb form in (9b) is homonymous in imperative and declarative mode. The declarative sentence could also by used in a generic meaning, equivalent to *A(ny) file can be opened by the* Open *command,* but in this case also the verb would have to be interpreted not as bound to the particular situation and one event of opening. We concentrate on the meanings concerning one particular action of opening throughout the example in (9).

(9c,d,e) all presume the salience of the Open command. (9d) presumes also a salient file, that is why we used the definite article in English. The contexts in which (9c) can be used are characterized by the question *What should we do by the* Open *command?*. It is also possible to use (9c) in a context characterized by the question *What should we do?* if it is somehow presumed that we are talking about using various commands (or various means or instruments) to do various things. In the latter type of context, the Open command in particular does not have to be salient. (9c) does not indicate specificity of the file.

For (9d), the appropriate contexts are characterized by the question *What should we do with the file by the* Open *command?* This is why this version is the most restricted.

(9e) presumes the salience of the opening action and of the *Open* command. The contexts for its use can be characterized by the question *What should we open by the* Open *command?*

(10) *Russian*

(a) *Откройте файл командой Open.*
Open-imp file-acc command-instr Open
Open a|the file by the Open command.

(b) *Файл откройте командой* Open.
File-acc *open*-imp *command*-instr Open.
Open the file by the Open *command.*

(c) *Командой* Open *откройте файл.*
command-instr Open *open*-imp *file*-acc
By the Open *command open a|the file.*

(d) °*Командой* Open *файл откройте.*
command-instr Open *file*-acc *open*-imp
By the Open *command open the file.*

(e) °*Откройте командой* Open *файл.*
open-imp *command*-instr Open *file*-acc
Open a file by the Open *command.*

(f) **Файл командой* Open *откройте.*
File-acc *command*-instr Open *open*-imp.
Open the file by the Open *command.*

(g) *Файл открывается командой* Open.
File-acc *open*-3-pers-refl *command*-instr Open.
A|the file is opened by the Open *command.*

(h) *Команда* Open *открывает файлы.*
command-nom Open *open*-3-pers *files*-acc-plur.
The Open *command opens* ***a*** *file.*

As well as in Czech, the word order of (10a) is neutral, while (10d-f) are impossible in a written text, same as the permutations in (4) above. (10b) and (10c) are possible with approximately the same acceptability conditions as in Czech. However, wide use of quasi-passive and passive constructions in Russian texts in this register suggests expressions of meanings corresponding to (10b) and (10c) as (10g) and (10h), respectively (cf. the sections on diathesis in this report). Generally the non-specific reading of the noun *file* can be emphasised in these examples by use of the plural number, while the use of specific determiners like *этот* (*this*) suggests definite reading, for example, in (10b) *Этот файл откройте командой* Open (cf. the discussion of determination in this report).

(11) *Bulgarian*

(a) *Отворете файла с командата* Open.
Open-imp *file*-acc *with command*-def-instr Open
Open the file by the Open *command.*

(b) *Файла отворете с командата* Open.
File-acc *open*-imp *with command*-def-instr Open.
Open the file by the Open *command.*

(c) *С командата* Open *отворете файла.*
With command-def-instr Open *open*-imp *file*-acc
By the Open *command open a file.*

(d) * *С командата* Open *файла отворете.*
With command-instr Open *file*-acc *open*-imp
By the Open *command open the file.*

(e) *Отворете с командата* Open *файла.*
Open-imp *with command*-instr Open-imp *file*-acc
Open a file by the Open *command.*

(f) * *Файла с командата* Open *отворете.*
File-acc *with command*-instr Open *open*-imp.
Open the file by the Open *command.*

The sentences (11d) and (11f) are ungrammatical in Bulgarian («poetic inversion» word order, see above). Normally, the sentence in (11a) is not neutral in Bulgarian, but is used as the sentences (11b,c,e) are used in the Czech context (9b,c,e), i.e. referring to opening a specific file. So in all the cases the noun is accompanied with a definite article (the problems concerning its full and short form in masculine are not discussed here). In fact the most natural variant according to its word order will be (11a), while the other variants may be used sometimes to stress more unusual theme.

The sentence

(9a') *Отворете файл с командата* Open.
Open a file by the Open *command.*

may also be used in the context of the Czech sentence (9a), but such a meaning (referring to any file) is supposed to be used rarely in instructional texts.

A more natural word order for (9b) in Bulgarian will be its passive form:

(9b') *Файлът се отваря с командата Open.*
The file is opened by the Open command

This set of examples in the three Slavonic languages corroborates the earlier observation that word order is more free in Czech and Russian than in Bulgarian. Czech and Russian allow the grammatical structure to be preserved and the constituents to be permuted according the communicative purpose in a given context. The contexts corresponding to some of the permutations may of course be less likely to occur in the AGILE texts. Bulgarian, on the other hand, seems to be close to English. The example in (11) suggests that like English, Bulgarian only allows at most one experiential element to precede the verb. This would be the Theme (see Section 2.7.2.1). In comparison to English, (11e) is interesting, because in this case Bulgarian allows a word order which in not possible in

English, i.e. *Open with command* Open *a file*. So the generalization appears to be that the order of elements after the verb is free, or at least more free, in Bulgarian than in English. We shall further explore the possibility that this order is determined by the communicative purpose, in the same way as in Czech and Russian.

Now we would like to consider another set of word order variations. It is also derived from the intermediate demonstrator text, and it modifies the sentence in (1) above. We do not include the highly marked versions with the verb ordered at the end.

(12) *Czech*

(a) *Otevřete panel* Styly multičáry *jednou z následujících metod.*
Open-imp *pane*-acc Multiline Styles *one*-instr *of following methods.*
Open the Multiline Styles *pane by one of the following methods.*

(b) *Panel* Styly multičáry *otevřete jednou z následujících metod*

(c) *Jednou z následujících metod otevřete panel* Styly multičáry

(d) *Otevřete jednou z následujících metod panel* Styly multičáry

The assumptions concerning salience are the same as in the previous example, namely that whatever precedes the verb is presumed to be salient in the context where the sentence is used.

The order in (12a) is the order used in the manual. However, the order in (12d) where Instrument precedes Goal is considered neutral in Czech (see Section 2.7.3.1). The motivation for «postponing» the Instrument here does not seem to be that the Goal should be considered contextually salient and the Instrument should not. It seems more that because the constituent expressing the Instrument is quite heavy, it is ordered at the end. So in this case, the difference in word order does not reflect a difference in meaning.

The difference between (12a) and (12c) is quite difficult to explain in terms of presumed salience in the context. Why should it be known to the addressee that several methods of doing something will be presented? This order could, however, be appropriate if as a matter of text planning strategy the decision is to chose methods as a departure point for the message, rather than the goal for which actions are carried out. This difference is also present in the following pair:

(13) *Czech*

(a) *Stisknutím* ENTER *kreslení ukončete.*
Pressing-istr ENTER-gen *drawing*-acc *end*-imp.
By pressing ENTER *end drawing*.

(b) *Kreslení ukončete stisknutím* ENTER.
Drawing-acc *end*-imp *pressing*-istr ENTER-gen
End drawing by pressing ENTER.

It does not seem appropriate to think that in one case it is the action that is more salient and in the other it is the Goal. Rather, this seems to be a text organization decision.

(14) *Russian*

(a) *Откройте панель* Multiline Styles *одним из следующих способов.*
Open-imp *pane*-acc Multiline Styles *one*-instr *from following methods.*
Open the Multiline Styles *pane by one of the following methods.*

(b) *Панель* Multiline Styles *откройте одним из следующих способов.*

(c) *Одним из следующих способов откройте панель* Multiline Styles.

(d) *Откройте одним из следующих способов панель* Multiline Styles.

Unlike Czech, the position of the reference to the method list at the end of (14a) in Russian is neutral and also satisfies the condition of a natural text deployment, so that this sentence is naturally linked to what follows in the list. Since the word order in Russian does not exhibit the same degree of flexibility of imperatives as in Czech, all the examples (12b-d) are unlikely to be found in a written text. In a dialogue this could be expressed using exophoric references to the situation:

(e) *Эту панель откройте так или так.*
This panel-acc *open*-imp *so or so.*

The literal Russian equivalents of the Czech sentences in (13) are unnatural in the same way as in (4c-e). The opposition of goal-centered vs. method-centered imperative sentences discussed for example (13) in Czech is expressed in Russian as follows, respectively:

(15) *Russian*

(a) *Нажмите* ENTER *для завершения рисования.* (goal-centered)
Press ENTER *for ending*-gen *drawing*-gen.
Press ENTER *to end drawing*.

(b) *Завершите рисование нажатием* ENTER. (method-centered)
End-imp *drawing*-acc *pressing*-instr ENTER.
end drawing by pressing ENTER.

Also, the «*for+nominalization*» construction is only marginally grammatical in Czech (see the discussion of expressing purpose in clause complexity in this report). So the nominalization is the most compact way of expression in Czech (a less compact way would be to use a finite subordinate clause).

In Bulgarian, the equivalents of the sentences suggested for Czech in (12) and (13) above are all grammatically well-formed. The conditions for acceptance discussed above hold for them as well. This observation corroborates the view that as long as at most only one experiential element appears before the verb in Bulgarian, the remaining elements can be ordered according to the communicative purpose in the given context.

(16) *Bulgarian*

(a) *Отворете диалоговия прозорец* Multiline style *с един от следните методи.*
Otevřete panel Styly multičáry jednou z následujících metod.

(b) *Диалоговия прозорец* Multiline style *отворете с един от следните методи.*
Panel Styly multičáry otevřete jednou z následujících metod

(c) *С един от следните методи отворете диалоговия прозорец* Multiline style
Jednou z následujících metod otevřete panel Styly multičáry

(d) *Отворете с един от следните методи диалоговия прозорец* Multiline style.
Otevřete jednou z následujících metod panel Styly multičáry

(17) *Bulgarian*

(a) *С натискане на ENTER чертаенето завършва.*
Stisknutím ENTER kreslení ukončete.

(b) *Чертаенето завършва с натискане на ENTER.*
Kreslení ukončete stisknutím ENTER.

All the examples discussed so far were formulated in the imperative form, so there was no explicit Actor. It is of course possible to reformulate them in declarative mode, with an explicit Actor, the user. Then there is even more possibilities for ordering. We shall not enumerate them all. However, note the following two possibilities where the verb is preceded by two participants of the process:

(18) *Czech*

(a) *Uživatel soubor otevře příkazem* Open.
User-nom *file*-acc *opens command*-instr Open.
User opens the file by the Open *command.*

(b) *Uživatel příkazem* Open *otevře soubor.*
User-nom *command*-instr Open *opens file*-acc
User opens a file by the Open *command.*

These both presume salience, i.e. contextual familiarity of the user. (18a) also presumes salience of a file, while (18b) presumes salience of the *Open* command. (18a) indicates a non-specific file. If a specific file were to be referred to, this would have to be made explicit by some form of determination, either by a Deictic element or an Epithet as below:

(c) *Uživatel příkazem* Open *otevře daný soubor.*
User-nom *command*-instr Open *opens given file*-acc
User opens the file by the Open *command.*

Similarly for Russian:

(19) *Russian*

(a) *Пользователь открывает файл командой* Open.
User-nom *opens* *file*-acc *command*-instr Open.
User opens the file by the Open *command.*

(b) *Командой* Open *пользователем открывается файл.*
command-instr Open *User*-inst *opens*-refl *file*-nom
User o*pens an existing file by the* Open *command.*

An existing file can be referred to as illustrated below in (c) where contrast is established with a creation of a new file:

(c) *Командой* Open *пользователем открывается существующий файл.*
command-instr Open *User*-inst *opens*-refl *existing* *file*-nom
User o*pens an existing file by the* Open *command.*

In Bulgarian, the equivalent of (b) is ungrammatical. Again, this fits with our hypothesis that Bulgarian only allows one experiential element to precede the verb:

(20) *Bulgarian*

(a) *Потребителят отваря файла с командата* Open.
User-nom *file*-acc *opens* *command*-instr Open.
User o*pens the file by the* Open *command.*

(b) * *Потребителят с командата* Open *отваря файла.*
User-nom with *command*-def-instr Open *opens* *file*-acc
User o*pens a file by the* Open *command.*

A reader who is familiar with the SFG framework can already see that the thematic structure and information structure can help us to capture the differences in meaning we have alluded to. For the sake of those readers who are not familiar with the SFG framework, the next section overviews the essential ideas.

2.7.2 Thematic Structure and Information Structure in SFG

In the SFG tradition, language is seen to serve three functions: the ideational, interpersonal and textual, which are all reflected in linguistic structure simultaneously. Halliday has explained in detail at various places that they are all semantically relevant and it is not necessary to argue that one function is more abstract than another. Subsequently, we can say that any particular word order of a sentence in a context results from an interplay between the multiple language functions.

2.7.2.1 Thematic Structure in SFG

According to Halliday (1985), clause as a message consists of a **Theme** combined with a **Rheme**, and in this configuration, «the Theme is the starting point for the message, it is the ground from which the clause is taking off» (p. 38). Halliday's SFG approach was initially inspired by the Praguian dichotomy of Theme and Rheme. Following the original terminology of the Prague School, Halliday says:

«The Theme is the element which serves as the point of departure of the message; it is that with which the message is concerned. The remainder of the message, the part in which the Theme is developed, is called in Prague School terminology the Rheme. As a message structure, therefore, a clause consists of a Theme accompanied by the Rheme; and the structure is expressed by the order – whatever is chosen as the Theme is put first.» (Halliday 1985, p. 37)

One should remark that the characterization of a Rheme as the remainder is slightly misleading, because it is in fact the Rheme that conveys the important part of the message. The Theme provides the grounding of the message in the context, and when this grounding is considered sufficiently obvious by the speaker, it can be omitted in the surface realisation of a sentence. Unlike the Theme, the Rheme cannot be omitted.

The following examples from (Halliday 1985, p.38) illustrate the Theme+Rheme structure:

the duke	has given my aunt that teapot
my aunt	has been given that teapot by the duke
that teapot	the duke has given to my aunt
Theme	*Rheme*

Halliday notes that as a general guide, the Theme can be identified as that element which comes in first position in the clause, although this is not how the Theme is defined. First position is the means whereby the function of Theme is realised, in the grammar of English and functionally similar languages. The following examples from (Halliday 1985, p. 39) show that the Theme is not necessarily a nominal group, but can also be an adverbial group or a prepositional phrase:

once upon a time	there were three bears
very carefully	she put him back on his feet again
for want of a nail	the shoe was lost
with sobs and tears	he sorted out those of the largest size
Theme	*Rheme*

Besides Themes consisting of just one structural element represented by just one unit, a Theme may, according to Halliday, also consist of two or more groups of phrases forming a single structural element, i.e. a group or phrase ‘complex’. Examples (p. 40):

the Walrus and the Carpenter	were walking close at hand
Tom, Tom, the piper’s son	stole a pig and away he run
from house to house	I went my way
on the ground or in the air	small creatures live and breathe
Theme	*Rheme*

Halliday also acknowledges the existence of a «special thematic resource whereby two or more elements are grouped together so that they form a single constituent of the Theme+Rheme structure. He calls it the 'thematic equative'. The examples he gives (p. 41) are as follows:

a.	what (the thing) the duke gave to my aunt	was that teapot
b.	the one who gave my aunt that teapot	was the duke
c.	the one the duke gave that teapot to	was my aunt
d	what the duke did with the teapot	was give it to my aunt
e.	how my aunt came by that teapot	was she was given it by the duke
	Theme	*Rheme*

Moreover, Halliday notes that certain elements have a special status in the TS in that they tend to be --or sometimes have to be -- thematic. He calls them **characteristic themes** and distinguishes between 'conjunctive (discourse) adjuncts', 'modal adjuncts', 'conjunctions' and 'relatives'. Conjunctive and modal elements are outside the experiential structure of the clause, as they have no status as participant, circumstance or process. However, since they are characteristic themes they can appear as part of the Theme. Halliday thus specifies the Theme as follows:

«[T]he Theme extends from the beginning of the clause up to (and including) the first element that has a function in transitivity. This element is called the 'topical Theme'; so we can say that the Theme of the clause consists of the topical Theme together with anything else that comes before it.» (Halliday 1985, p. 52)

A Theme thus always contains one and only one experiential element, which may be preceded by elements which are textual and/or interpersonal in function; if so, these are also part of the Theme. Such Themes that consist of a topical Theme and one or more characteristic themes are called **mutiple themes** by Halliday. The typical ordering is *textual^interpersonal^experiential*.

The question which element of the clause is typically chosen as the Theme depends on the choice of Mood. The usual pattern according to Halliday can be summarized as follows:

Mood of clause	*Typical ('unmarked') Theme*
declarative	nominal group functioning as Subject
interrogative: yes/no	first word (finite operator) of verbal group, plus nominal group functioning as Subject
interrogative: WH-	nominal group, adverbial group or prepositional phrase functioning as interrogative (WH-) element
imperative: 'you'	verbal group functioning as Predicator, plus preceding *don't* if negative
imperative: 'you and me'	*let's*, plus preceding *don't* if negative
exclamative	nominal or adverbial group functioning as exclamative (WH-) element

2.7.2.2 Information Structure in SFG

As noted earlier, (Halliday 1985) distinguishes between the thematic structure of a clause and what he calls the information structure. The latter is the distinction between **Given** and **New** within an information unit: the speaker presents information to the listener as recoverable (Given) or not recoverable (New). Givenness and newness are closely related to intonation patterning in Halliday's approach. The New is marked by prominence, and the (optionally present) Given typically precedes the New.

The thematic structure and information structure are closely related but not the same according to Halliday. The Theme is what the speaker chooses to take as the point of departure, the Given is what the speaker believes the listener already knows or has accessible. As for the relation between the two dichotomies, Halliday says:

«Other things being equal, a speaker will choose the Theme from within what is Given and locate the focus, the climax of the New, somewhere in the Rheme.» (Halliday 1985, p. 299)

«[T]heme will be associated with the 'given' and rheme with the 'new' unless there is good reason for choosing some other alignment» (Halliday 1970, p. 162).

In English, according to Halliday, information structure is expressed by intonation.

2.7.2.3 Word Order in SFG

The notion of Theme as Halliday formulates it, tells us a number of things about «the first» position in the clause, but it does not tell us much about the word order of «the rest» of the clause, as far as we can see. Presumably, Halliday leaves this to be decided by the grammatical structure. However, as we noted before, in languages with a high degree of free word order the grammar is not very strict about the placement of the groups and phrases within the clause. The examples we discussed above showed that ordering in our languages is to a great extent determined by what is presumed to be salient in the context, or recoverable by the addressee. This means ordering reflects information structure.

Let us now consider the question whether the thematic structure and information structure according to Halliday's proposal (for English) are sufficient to capture the word order of Czech sentences. We'll first have a look at the possibility to generate the Czech sentences in (21).

The word order in (21a) is easy to describe, because it is the same as the normal word order in the English sentence. The Actor is conflated with the Subject, the Finite follows the Subject, the Goal follows the verb and the Instrument appears last.

The word order in (21b) is also easy to describe using the apparatus available in SFG. The only difference from (21b) is that this time the Instrument is thematized.

But when we look at (21c) and (21d), we can see that the word orders do not correspond to any grammatical structure of English, and they cannot be obtained only by thematization. A possible way of describing the word order in (21c) is that 'file' is Given, and 'user' is either Given or Theme or both. Then, using the principle that Given comes before New, the displayed word order can be obtained.

Likewise, in order to generate (21d), we consider 'the *Open* command' as Given, and 'user' as either Given or Theme or both.

(21) *Czech*

(a) *Uživatel otevře soubor příkazem* Open.
User-nom *opens file*-acc *command*-instr Open.
User o*pens the file by the* Open *command.*

(b) *Příkazem* Open *uživatel otevře soubor.*
C*ommand*-instr Open.*user*-nom *opens file*-acc
B*y the* Open *command, user* o*pens the file.*

(c) *Uživatel soubor otevře příkazem* Open.
User-nom *file*-acc *opens command*-instr Open.
User o*pens the file by the* Open *command.*

(d) *Uživatel příkazem* Open *otevře soubor.*
User-nom *command*-instr Open *opens file*-acc
User o*pens a file by the* Open *command.*

Now, how about the ordering of the New elements? Are there any rules to follow? Consider the following examples:

(22) *Czech*

(a) *Objeví se dialogové okno*
Appears-refl *dialogue window*-nom.
A dialogue box appears.

(b) *Dialogové okno se objeví.*
D*ialogue window*-nom *appears*-refl.
The *dialogue box appears.*

(23) *Russian*

(a) *На экране появится диалоговое окно.*
On screen-loc *appear*-refl *dialogue window*-nom.
A dialogue box appears.

(b) *Диалоговое окно исчезнет с экрана*
Dialogue window-nom *disppears*-refl *from screen*-gen.
The *dialogue box disappears.*

(24) *Bulgarian*

(a) *Появява се диалоговия прозорец.*
Appear refl *dialogue window*-nom.
A dialogue box appears.

(b) *Диалоговият прозорец се появява (на екрана).*
D*ialogue window*-nom refl *appears*-refl.
The *dialogue box appears.*

(c) *Появява се диалоговият прозорец.*
Appear refl *dialogue*-def *window*-nom.
A dialogue box appears.

In all the languages, as well as in English, the 'dialogue box' is the Subject in both (a) and (b). Clearly, if it precedes the verb and is not accompanied by a deictic in Czech and Russian, it refers to a Given, familiar object (it has a definite article in Bulgarian, cf. 24b). The (24c) version in Bulgarian forms a minimal pair with (24a) with respect to the specificity of 'dialogue box'. So with respect to specificity (24b) and (24c) both correspond to the Czech (22b), where 'dialogue box' is interpreted as familiar.

So when we want both the process and the Actor to be understood as New, we should order them as in (a). In this way, attention is directed at the whole event and the messages culminates with what appears. In the opposite order, attention would be directed towards the process itself, but such information structure is not appropriate a context where neither the process not the Actor are familiar.

There are more general issues that concern the ordering of New elements. These issues have been studied in detail in the Praguian FGD framework. We shall overview the most essential ideas in the following section, and propose their incorporation into the SFG view on information structure and word order.

2.7.3 Information Structure in FGD[11]

The framework of Functional Generative Description (FGD) has been developed within the Prague School of Linguistics as a functional approach to the description of language. FGD works with a notion of information structure consisting of one dichotomy into **Topic** and **Focus**. The notion of information structure in FGD is referred to as **Topic-Focus Articulation** (TFA). Even though, as mentioned earlier in a quote from Halliday's work, the two approaches share the same origin, their later versions diverge in some respects. The obvious difference is that FGD works with one structure and not with two. More subtle differences would concern the division of particular sentences into the Topic and Focus part in comparison to the division into the Theme and Rheme and/or Given and New in Halliday. However, it is not relevant for our current purposes to go into the details of such a comparison. We will only describe those notions that we would like to incorporate into the SFG treatment of word order for the sake of handling Slavonic languages.

The notions we need to introduce are the following:

- dependency relationships between units of meaning
- communicative dynamism
- contextual boundness (**CB**) versus non-boundness (**NB**) for each lexical item[12]

The representation of the meaning of a sentence in FGD has the form of a **dependency structure**. Its units correspond to autosemantic lexical items. The main structure arises from how the various units are related to one another via dependency relations. In such a relation, one unit (the dependent node) is said to modify another unit (the governing node, or head) via a specific kind of dependency relation. A set of dependency relations is assumed, among which there are ACTor, PATient, ADDressee, EFFect, ORIgin, LOCation, DIRection,

[11] The text in this section uses (Kruijff-Korbayová 1998, Kruijff 1998, Partee et al. 1998 and Sgall et al. 1986) as the main sources.

[12] A suitable operational test for this distinction can be based on question-answer pairs or on negative and other continuations.

TEMPoral, APPurtenance, etc. For a detailed discussion of the kinds of dependency relations see (Sgall et al. 1986, Petkevič 1995).

2.7.3.1 Systemic Ordering and Communicative Dynamism

Given the set of kinds of dependency relations discerned for a given language, one can define an ordering in which complementations, i.e. «arguments» and «adjuncts», occur in the deep word order in an unmarked case. This ordering is called the **systemic ordering** (SO) in FGD. It is a total ordering over all possible dependency relations. It may differ from one language to another, but is considered universal within a given language. Systemic ordering in Czech has been discussed in most detail in (Sgall et al. 1980, 1986), (Hajičová and Sgall 1987), (Petkevič 1995); for English see (Hajičová et al. 1993) and for German in (Sgall et al, 1995). Similar ideas for Japanese have been developed for instance by Kuno. The systemic ordering for Czech formulated in terms of the FGD types of complementations is as follows (Sgall et al. 1986):[13]

Actor < «from when» < «when» < «how long» < «till when» < «for how long» < Cause < Aim < Manner < Accompaniment < Locative < Means < Difference < which way < Addressee < Origin < from where < Objective < Directional < Effect < Comparison < Appurtenance < General Relation

We expect the SO for there the main types of complementations in Russian and Bulgarian is the same as the Czech one, with on exception: when analyzing the examples in this report, we have observed that in Russian, the order of Goal before Instrument is neutral. This means, in the FGD terminology that Means should be placed somewhere before Objective in the systemic ordering. Further research ad experiments are needed verify this hypothesis and other possible differences between the SO in Czech, Russian and Bulgarian. Before we incorporate the systemic ordering into the KPML grammar, we have to map the FGD types of complementations to the SFG ones. For the Participants according to SFG, i.e. $Actor_{SFG}$ and $Goal_{SFG}$, this is easy, they correspond to the $Actor_{FGD}$ and $Objective_{FGD}$ (often also called $Patient_{FGD}$). Also the various types of Spatial and Temporal locatives are easy to compare. The explanation of the rest disserves a further insight before we can say which SFG Circumstatials are involved. Here we provide some initial hints:

- Aim_{FGD} corresponds to «for the sake of» in English»
- $Manner_{FGD}$ corresponds to «how»
- $Accompaniment_{FGD}$ corresponds to «with»
- $Means_{FGD}$ corresponds to $Instrument_{SFG}$, as in *write by a pen*)
- $Difference_{FGD}$ is exhibited for instance in *she moved <u>one inch away</u>*
- $Adressee_{FGD}$ corresponds to «to (for) whom», it is understood narrower than $Beneficiary_{SFG}$
- $Origin_{FGD}$ corresponds to «from (out of) what» as in *make a piece of art <u>from an odd piece of wire</u>* (see the discussion in the section on transitivity in Czech in this report)
  - $Effect_{FGD}$ corresponds to «appoint someone something» or «make something into something», so it is a kind of result
  - $Comparison_{FGD}$ is exhibited in «as big as» or «bigger than»
  - $Appurtenance_{FGD}$ is exhibited in *the style <u>of</u> the line* or *the element <u>of</u> the line*
  - $Cause_{FGD}$ corresponds to «for what reason»
  - $General\text{-}Relation_{FGD}$ corresponds to a restrictive or non-restrictive attribute

[13] We leave out some types of complementations which are not encountered in the AGILE corpus.

The order of nodes in a dependency structure corresponding to a sentence under a linear projection is called **communicative dynamism** (CD) in FGD. This notion originates from the works of Firbas (Firbas, 1992). The degrees of CD in a given sentence may but do not have to agree with the SO. The relation between SO and CD is often illustrated using the asymmetry in the following two English sentences:[14,15]

(25)

(a) *They flew from Chicago to* BOSTON.

(b) *They flew to Boston from* CHICAGO.

While (a) is unmarked, (b) can only be uttered felicitously if Boston is salient in the (verbal or situational) context. This asymmetry can be accounted for on the basis of SO: if (i) in English, direction-from precedes direction-to under SO, and if (ii), in general, a departure from SO marks what is displaced to the left as less communicatively dynamic.

As the examples below show, a similar relationship can be found with other pairs of dependency relations. They also illustrate the influence of the placement of the intonation center, as another parameter in determination of CD, which appears especially important in languages like English, where deviations from the standard grammatically rather fixed word order are accompanied by marked intonation pattern. Czech does not use intonation for this purpose, because the grammar restricts word order to a much lesser degree.

(26)

(a) *They went by car to a* RIVER.

(b) *They went to a river by* CAR.

(27)

(a) *Jim dug a ditch with a* HOE.

(b) *With a hoe, Jim dug a* DITCH.

(c) *Jim dug a* DITCH *with a hoe.*

(28)

(a) *Ron cannot sleep quietly in a* HOTEL.

(b) *In a hotel Ron cannot sleep* QUIETLY.

Each of the (a) and (b) examples has an unmarked intonation. In the (a) examples, CD agrees with SO, whereas in the (b) examples the CD of the last two complementations in each sentence is reverted with respect to SO. This disagreement between CD and SO

[14] See (Partee et al. 1998) for the most recent discussion.

[15] We use the typographical notation of marking the sentence or clause intonation center by SMALL CAPITALS. We make the assumption usual in FGD that the intonation center regularly has the shape of Pierrehumbert's H*L and that its prototypical position is at the last clause part that may be stressed.

appears in the (c) examples as well, due to the marked intonation pattern, in spite of the fact that the surface word order agrees with SO, like in the (a) examples.

Examples of this kind have been analyzed with several series of psycholinguistic tests for Czech (Sgall et al. 1980), and recently also for German and in a lesser extent for English (cf. (Partee et al. 1998) for references).

2.7.3.2 Contextual Boundness

The interplay between SO and CD is closely related to the distinction between **contextual boundness** (**CB**) and **non-boundness** (**NB**) for each node in a TR. A **CB** lexical item is assumed to convey some content that bears a **contextual relationship** to the discourse context. Such an item may refer to an entity already explicitly referred to elsewhere in the discourse, or to an entity «implicitly evoked» in the discourse context. In the framework of FGD, one says that the referent of a **CB** linguistic item is easily available in the current discourse context (see Hajičová 1993 for a summarizing discussion).[16]

An operational test for the identification of the value of the CB feature using the question test has been formulated in (Sgall et al. 1986). It can be summarized as follows:

- weak (and zero) pronouns are always **CB**
- those items in an answer which reproduce expressions present in the question representing a context in which the tested sentence can appropriately be uttered are **CB**
- other items are **NB**

Let us illustrate the application of the test using the example we discussed in the section on word order in SFG. We repeat it below as (29).

(29) *Czech*

(a) *Uživatel*$^{cb/nb}$ *otevře*$^{cb/nb}$ *soubor*^{nb} *příkazem*^{nb} Open^{nb}.
User-nom *opens file*-acc *command*-instr Open.
User o*pens the file by the* Open *command.*

(b) *Příkazem*^{cb} Open^{cb} *uživatel*^{cb} *otevře*$^{cb/nb}$ *soubor*^{nb}.
C*ommand*-instr Open.*user*-nom *opens file*-acc
B*y the* Open *command, user* o*pens the file.*

(c) *Uživatel*^{cb} *soubor*^{cb} *otevře*$^{cb/nb}$ *příkazem*^{nb} Open^{nb}.
User-nom *file*-acc *opens command*-instr Open.
User o*pens the file by the* Open *command.*

(d) *Uživatel*^{cb} *příkazem*^{cb} Open *otevře*$^{cb/nb}$ *soubor*^{nb}.
User-nom *command*-instr Open *opens file*-acc
User o*pens a file by the* Open *command.*

The CB/NB marking in the example is explained by the fact that with unmarked intonation, the questions that can be answered by the individual sentences are as follows:

---

[16] A contextually bound item can also refer deictically, and be thus related to the extra-linguistic context of the discourse. However, I disregard deictic reference for the present purposes.

(30)

(a) *What happens?*
or *What does the user do?*

(b) *What does the user do by the Open command?*
or What does the user open by the Open command?

(c) *What does the user do with the file?*
or *How does the user open the file?*

(d) *What does the user do by the* Open *command?*
or *What does the user open by the Open command?*

Contextual boundness is not identical to previous mentioning in the (verbal) co-text. Sgall et al. state that it appears that in the structure of language the cases in which a CB item is directly known from the co-text are not distinguished from those in which they are adduced as (if) readily available in the speaker's memory.

It appears to be the that case the distinction between contextual boundness and non-boundness in FGD is very similar to that between Given and New proposed in SFG. It seems however that the CB/NB has been studied and discussed in more detailed at least by the Praguian scholars. Therefore, for the purposes of handling word order in Slavonic languages in the SFG framework, we propose to replace the Given vs. New distinction by the CB/NB distinction as explained within FGD.

2.7.3.3 Topic-Focus Articulation

Topic-Focus Articulation (TFA) of a sentence represents the FGD's account of information structure. At the level of a clause or simplex sentence, TFA corresponds to a division into the **Topic part** and the **Focus part**. In complex sentences, TFA is conceived of as a **recursive structure**.

The notions now called «topic'» and «focus» in FGD can be traced back to (Weil 1887). Weil's work was resumed by several German linguists in the decades around the turn of the millennium, and subsequently in the Prague Circle a systematic attention was paid to issues of information structure. Mathesius recognized that the distinction between topic and focus was important to problems ranging from intonation to word order -- issues central to the description of natural language (Mathesius 1936, Mathesius 1939). Within FGD, the theory of TFA was further developed by Sgall and his collaborators, in particular Hajičová (cf. (Hajičová 1993) for a summary and references).

In the FGD approach, the dichotomy of Topic and Focus is derived from (based on) the structural notions of contextual boundness and non-boundness, rather than being conceived of as primary like in most other theories of information structure. The following abstract definition of topic and focus as to their relation to CB and NB is given in (Sgall et al. 1986, p. 216):

Given a tectogrammatical representation of a sentence:

(i) the main verb belongs to the focus if it is NB, and to the topic if it is CB;

(ii) the NB nodes depending on the main verb belong to the focus, and so do all nodes (transitively) subordinated to them;

(iii) if some of the elements of the tectogrammatical representation belong to its focus according to either of the above points, then every CB daughter of the main verb together with all nodes (transitively) subordinated to it belong to the topic;

(iv) if no node of the tectogrammatical representation fulfills the first two points above, then the focus may be more deeply embedded; special rules for the determination of Focus are applied in these cases

The least communicatively dynamic item in the Topic is called **topic proper**[17], and the most communicatively dynamic item in the Focus is called **focus proper**. The Topic may be empty, e.g. in discourse-initial sentences.

In the primary cases of TFA, the Topic-Focus boundary immediately precedes or follows the main verb-node in the TR of a sentence.

In the secondary cases of TFA with the Focus being deeper embedded, the verb and (at least) the nodes directly dependent on it belong to the Topic, and the boundary is deeper embedded.

(10) In the general case, neither Topic nor Focus is a single item (at any level).

It also follows from the FGD definition of TFA that the view that the Topic consists of the CB items and the Focus consists of the NB ones, is too simplistic and as such only holds in simple cases. An NB node can belong to the Topic, if its governing node belongs to the Topic, and a CB node can belong to the Focus if its governing node belongs to the Focus.

2.7.3.4 TFA and Word Order

As we noted earlier, information structure is considered the main determining factor for word order in Czech and other languages with a high degree of free word order. In fact, also in other languages, e.g. English, word order can often follow information structure. For instance, the use of passive instead of active voice can be in many cases, especially in written texts, explained by the need for the order Goal/Subject<Actor/Object rather than Actor/Subject<Goal/Subject. However, the word order is much more rigid in English, since the grammatical structure does not allow such a degree of freedom in the placement of groups and phrases within a clause, as it is in Czech and most other Slavonic languages. We shall now summarize the FGD claims concerning word order.

1. The main principle of word order in Czech is that the Topic precedes the Focus. Since the Topic may be empty (esp. in discourse-initial sentences) or it may be deleted on the surface due to ellipsis, it is possible that the surface form of some sentences only consists of the realisations of elements belonging to the Focus.
2. In the primary cases when the Topic consists of the CB elements, and the Focus of the NB ones, one can say that the CB elements precede the NB elements.

[17] It is worth noticing that in many cases the topic proper according to FGD corresponds to the Theme according to SFG.

3. A more general formulation of principle 2 uses the degrees of CD: in primary cases, CD and the surface word order correspond to each other quite closely, at least within clauses. So, the ordering from left to right in the surface realisation corresponds to the increasing degrees of CD. There are the following exceptions to this principle in Czech:
   (i) *clitics*: in the surface word order they have to be placed in the so-called **Wackernagel's position**, characterized as the second position in a clause
   (ii) *the main verb*: its preferred default (unmarked) placement is after, but not necessarily immediately following, the surface Subject, if there is one, even in cases where the verb precedes all its complementations under CD
4. Last but not least, FGD assumes that on the preferred default reading an adjective is more communicatively dynamic than its head noun, even though in the surface word order it is usually placed before it (the other way round sounds archaic in most cases). On the other hand, adjuncts constituted by proper names (of people assumed by the speaker to be known and recoverable by the hearer) are considered less dynamic than their head nouns

As the starting point for specifying the principles of word ordering in the context of AGILE, we propose to combine this FGD-based strategy which reflects information structure with the possibility of thematization in the usual SFG spirit.. For Czech and Russian, we need to allow for more freedom in word order (i.e., a looser relation between ordering and grammatical structure) than in Bulgarian. Namely, in Bulgarian there seems to be the restriction proposed in Section 2.7.1 that only one experiential element may preced the verb. Other restrictions or preferences can be included in the grammars of the specific languages, for instance the Russian dislike for a nominalization as the first element of a clause (cf. Section 2.7.1).

We propose to preserve the SFG notion of thematic structure with Theme as being the first experiential element in the clause and being determined by text organization. This conception appears useful especially in Bulgarian, as noted earlier. But also in Czech and Russian we find it useful in order to account for text structuring concerns across sentences within connected spans of texts. For instance, the decision what to chose as a «point of departure» can be motivated by a particular style chosen for the text, in which case it is not needed to look for a motivation for a particular ordering based on information structure.

2.7.4 An Abstract Algorithm for Word Order

In this section we present an abstract algorithm for achieving an ordering of clause constituents that reflects their informative status relative to the sentence as a whole in a given context. We commence by discussing, in brief, the assumptions that the abstract algorithm is founded on, and the basic ideas behind the functioning of the algorithm. Thereafter, we present the algorithm itself as a conceptual design.

2.7.4.1 Assumptions and Basic Ideas

In the preceding sections we discussed word order and its relation to information structure in terms of Halliday's **Theme**, and the Praguian notions of **contextual boundness** (**CB**) and **contextual nonboundness** (**NB**) as well as the notions of **topic** and **focus** that are derived from the former two (CB, NB). For the algorithm, we assume that we can combine these

notions, giving rise to the following template for sentential structure from the viewpoint of informativity:

α	β	γ	δ	η
Theme	[Clitics]	CB elements	Verb	NB elements

Figure 19 - Template sentence structure regarding informativity

The α-element corresponds to Halliday's Theme. There are several reasons for combining the notion of Theme with to the Praguian notions of CB and NB (Topic-Focus Articulation) into an account of informativity, most of which we will discuss in more detail in the workpackage on Text Structuring (WP5). An important reason that regards the issue of word order from a purely structural viewpoint is the following.

In Slavonic languages, clitics must appear in what is called the Wackernagel position. The Wackernagel position can be characterized as the place between the first and second element of a clause. Naturally, this leaves to be defined what «element» means. It is easy to show that «first element» does not equate to «first constituent» in the sense of constituency-based grammar, since the element can be of arbitrary complexity[18]. A fortunate characteristic of Halliday's Theme is that it does delineate a clausal constituent that corresponds exactly to what should be regarded as «the first element» upon which the Wackernagel position follows.

The template structure identifies the Wackernagel position as β. It is in this position that we ought to place clitics like the reflexives that occur in the AGILE corpora.

After the Theme and (zero or more) clitics, we have the CB elements, the verb, and the NB elements. We recall from our earlier discussion on the Praguian notions that (a) there need not be any CB elements in a sentence, and that (b) the NB elements are ordered according to the language's **systemic ordering**. There are no such ordering constraints on the CB elements From a generative point of view, there ordering is determined by their relative **communicative dynamism**[19]**,** though see below**.**

For the purpose of the algorithm we assume that we can specify in an SPL whether an element is CB or NB. Such determination can be done in a manner similar to the way Theme is specified: Just like `:theme <id>` we could have `:cb <idx>` and `:nb <idy>`. Naturally an element cannot be both CB and NB, but it should be possible to have more elements specified as being CB or NB, (contrary to `:theme`). To be able to determine the linear order in which the CB elements should appear in the surface realisation, we assume that there is an additional field in the SPL to specify the relative communicative dynamism of the CB elements. For example, we could have `:commdynam` Δ, with Δ a list of identifiers specified as being CB (excluding the verb).

[18] Dependency grammar has clearly the edge over constituency grammars on this point, since it can be shown that «first element» *does* correspond to the first element of a dependency structure (be that a head, or a dependent). In DG, the Wackernagel position is thus simply defined as the place between the first and second element of a dependency structure.

[19] Whereas from the viewpoint of analysis one would go by the linearity of the clause in which the elements appear.

With Δ given, we take the following approach to ordering CB elements:
Given a list Φ of CB elements to be ordered,and Δ as an order over them:

1. If the verb is CB, then place the verb at the end of the CB-block γ, occupying δ.
   If Δ is not empty, apply Δ to the elements of Φ therewith making up the CB-block γ;
   Otherwise, form γ from the elements Φ in the order as they appear in Φ.
2. Apply whatever other ordering constraints apply.

2.7.4.2 The Ordering Algorithm

Below we present a concept of the abstract ordering algorithm. It is important to realise that the algorithm operates on the constituents *at the level of the clause*. Ordering within groups is currently assumed to be taken care of by the grammar, for example, the ordering of an adjective before the noun it modifies, even though ideally this the ordering of sub-clause constituents should also follow information structure.

```
Given a list Γ of ordering constraints imposed by the grammar,
Given a list L1 of constituents that need to be ordered,
Given a list Δ giving communicative dynamism of cb constituents,

    create empty lists LC and LN

    % LC is going to contain cb constituents (γ), LN nb constituents (η)

    repeat for each element E in L1
                if E is CB,
                      then add E into LC,
                      else add E into LN.
    if the verb is CB,
          then  Order the verb at the end of LC
                Order the remainder according to Δ
          else  Order all elements in LC according to Δ
    % Thus, if e_i < e_j in Δ, then e_i < e_j in LC except for the verb.
    if Γ ≠ ∅ then
          Order elements in L1 using ordering constraints in Γ
    % These constraints may, for example, involve ordering
    %  «heavy» constituents at the end of the clause.
```

Figure 20 - Abstract algorithm for word order

2.7.5 Conclusions

We have discussed some aspects of word order in Czech, Bulgarian and Russian. We compared examples with word order variations and discussed the similarities and differences among the three languages as far as grammaticality of the various versions in concerned and also with regard to their acceptability in different contexts.

As expected, we did encounter differences. Bulgarian is a language with a more restricted word order than Czech and Russian. And also among the latter two, some small differences have been identified when analyzing our simple examples.

One conclusion we have reached after scrutiny of the corpus is that whereas Czech and Russian can form grammatical sentences with more than one constituent expressing an experiential element preceding the verb, this seems impossible in Bulgarian. Like English, Bulgarian appears to allow for at most one experiential element preceding the verb. We can account for this using the SFG notion of Theme. We have not studied the issue whether

Bulgarian has some equivalent of the thematic equatives in English which would allow for more than one experiential element to precede the verb.

On the other hand, we found out that Bulgarian does allow more word order freedom than English with respect to the elements following the verb, at least in some configurations. Our conjecture is that the Bulgarian word order follows information structure in this case.

In Czech and Russian, as expected, there is even more ordering flexibility. In most cases information structure can be reflected simply be the appropriate ordering of elements regardless of the grammatical structure if the sentence. However, we did encounter some cases, especially in Russian, where simple permutation of elements did not yield satisfactory results. In particular, we encountered a problem with an attempt to order nominalization expressing means at the beginning of a sentence. In some cases Russian also seems to prefer a mediopassive construction while Czech uses a permutation in active voice.

The issues of word order are complex and diserve further study. At this stage of the project, we have formulated an algorithm realizing word ordering which applies at the level of a clause and reflects information structure by the corresponding ordering of the elements in the clause structure, unless the grammar has already imposed other ordering constraints. This proviso allows us to take reflect the observations stated above.

We use allow for the use of the SFG notion of Theme to achieve certain effects of text structuring. Also, the notion of Theme enables us to restrict the number of elements preceding the verb in Bulgarian. And another benefit of using Theme is that we define the placement of clitics relatively to it.

When the grammar does not determine any ordering, we use the notions of systemic ordering and communicative dynamism which we adopt from the framework of FGD. We allow for the discrimination between contextually bound and non-bound elements in the semantics of a sentence. Essentially, we order them separately following systemic ordering, placing the verb in between these two «lists».

The approach to word order we propose here is satisfactory to the extent that it is satisfactory in general to consider word order as a «second order» phenomenon, i.e. as an ordering that is applied to constituents in a structure that has already been generated regardless of information structure. This is the perspective taken in SFG, namely that textual phenomena are considered second order with respect to ideational and interpersonal.

However, a number of examples we discussed above suggest that this view is not satisfactory. Information structure and textual phenomena should be taken into account simultaneously with the ideational and interpersonal perspectives. In such a way we could capture the interplay of these factors in cases when a certain ordering which is required from the perspective of information structure motivates a certain choice of grammatical structure. For instance, we would be able to model the situation when the textually motivated ordering Goal< Actor results in generating a passive construction.

2.8 Determination

In the course of natural language communication, a speaker can choose various ways of expressing the experience of phenomena of the real world (or some picture of it). This function of language is reflected by the **experiential dimension** in SFG, which reflects meaning as organization of experience. One of the aspects of the experiential dimension is **reference** to objects and processes. In a broad perspective, referring is a function pertaining

to linguistic expressions of any complexity and granularity. Our focus in this part of the report is however narrower, as we concentrate on nominal groups, and in particular on determination as one the aspects involved in nominal reference.

The experiential structure of a nominal group as a whole has the function of specifying a class of things and some category of membership within this class. The element expressing the class is referred to by the functional label Thing. Categorization within the class is typically expressed by one or more of the functional elements accompanying the Thing. Determination is that aspect of the experiential structure which reflects an indication whether the class of things referred to by the nominal group is specific (unique) or non-specific (non-unique).

The experiential structure of a nominal group according to (Halliday 1994) can be depicted by the following diagram:

Deictic	Numerative	Epithets	Classifier	Thing	Qualifier

The components of the experiential structure in this view are distinguished by their functions. The Function of indicating whether a specific or non-specific subset of Thing is involved, is ascribed to the Deictic element. We can thus say that the Deictic element within the nominagl group carries the Function of Determination. This can be formalized by specifying that a Deictic element needs to be inserted when a nominal group is being generated (see Figure 21 for the corresponding part in the grammar).

```
NOMINAL
{nominal-relative;nominal-interrogative;
   nominal-specific; nominal-non-specific} →
 [nominal] (+Deictic)
```

Figure 21 Insertion of a deictic element

When specified in this way, a Deictic element is inserted in every nominal group, and it is inserted as a separate lexical item. We shall see below that such specification might need to be reconsidered for Slavonic languages.

This part of the report is concerned with the essential types of determination in Czech, Russian and Bulgarian in comparison to English, and with the various possibilities of realisation if the Deictic element in these languages in general, as well as with respect to instructional texts.

Our account is geared towards a description of determination for our three languages in terms of the SFG framework. No such description is readily available in the linguistic literature on Slavonic languages, so we recourse to some available approaches, and try to transform their findings into SFG terms. In this effort, we take the Czech language as a starting point for the investigation of similarities and differences with respect to English, and then we extend the comparison to also Russian and Bulgarian. We present our discussion in as much integrated fashion as possible, in accordance with the overall aim of stressing the cross-linguistic aspects of study and description adopted within this project.

The structure of this part of the report is as follows. In Section 2.8.1 we introduce the terminology we are working with. First we overview the SFG English-based account of determination offered in (Halliday 1994). Then we present the essential points of an account

of determination in Czech according to (Hlavsa 1975), which also draws on earlier accounts of determination in Czech and in other Slavonic languages.

Sections 2.8.3 and 2.8.4 discuss specific and non-specific determination in detail, respectively. In each of them, we compare Hlavsa's and Hallidays' views and offer a merged account in SFG terms first for Czech, and then by means of a contrastive extension also for Russian and Bulgarian. These detailed discussions apply to determination in general, but we also accompany them by evidence drawn from our corpus of instructional texts from the AutoCad manual. We summarize each detailed discussion by presenting a formal specification in terms of systemic networks.

In Section 0 we conclude our current account of determination and summarize our observations concerning the comparison between the three Slavonic languages, as well as their comparison to English.

2.8.1 Introduction of Terminology

In the SFG framework explained in (Halliday 1994), the deictic element indicates whether or not some specific subset of the Thing is intended; it is either specific or non-specific. The English specific deictics are classified into (i) demonstrative versus possessive, and (ii) determinative versus interrogative. The non-specific deictic items are classified into (i) total versus partial and (ii) singular, non-singular and unmarked. The partial deictics are further subdivided into selective and non-selective, the total ones into positive and negative. The items functioning as deictics in English are summarized in tables in Figure 22 and Figure 23, which are reproduced from (Halliday 1994 p. 181 and p. 182) and slightly elaborated on the basis of (Halliday and Hasan 1976).

Halliday's classification also contains the function POSTDEICTIC, or $DEICTIC_2$. Halliday characterizes it as a second deictic element within a nominal group which adds further to the identification of the subset of 'thing' in question by referring to its fame or familiarity, status in the text or similarity/dissimilarity to some other designated subject. Examples are *same, other, different, identical, expected, necessary, original aforementioned*, etc. It appears that the post-deictic elements can serve a variety of functions, including comparative reference, textual metafunction etc.

<table>
<tr><td colspan="3" rowspan="2">Specific deictic items</td><td colspan="3">Determinative</td><td rowspan="2">Interrogative</td></tr>
<tr><td colspan="3">Proximity</td></tr>
<tr><td></td><td></td><td></td><td>near</td><td>far</td><td>neutral</td><td></td></tr>
<tr><td rowspan="3">Demonstrative</td><td rowspan="2">selective</td><td>singular/mass</td><td>this</td><td>that</td><td rowspan="2"></td><td rowspan="2">which, what</td></tr>
<tr><td>plural</td><td>these</td><td>those</td></tr>
<tr><td colspan="2">non-selective</td><td></td><td></td><td>the</td><td>whichever, whatever</td></tr>
<tr><td rowspan="2">Possessive</td><td>pronominal</td><td>person</td><td colspan="3">my your our
his her its
their
one's</td><td>whose(ever)</td></tr>
<tr><td colspan="2">non-pronominal</td><td colspan="3">[John's], [my father's] etc.</td><td>[which person's] etc.</td></tr>
</table>

Figure 22 Items functioning as specific Deictic in English

Non-Specific deictic items			**Number**				
			singular		non-singular		unmarked
					dual	mass/plural	
Total	positive	individual	*each*		*both*	*all*	
		collective	*every*				
	negative			*neither (not either)*			*no (not any)*
Partial	selective		*one*	*either*			*some [s^m]* *any*
	non-selective		*a(n)*			*some [sm]*	

Figure 23 Items functioning as non-specific Deictic in English

When looking for a useful source of analysis of deictic items in Czech, the most elaborated and most recent is the work of Hlavsa (Hlavsa 1975) on denotation in contemporary Czech language, an account also reflected in the Academic grammar (Daneš et al. 1987).

Hlavsa uses the term **designation** for the relation between an expression, and a class of objects that the expression designates, and the term **denotation** for the relation between a denotationally used expression and an object from the class of objects designated by the expression. One can write *p(x)* for designation (where *x* is a variable for the objects designated by expression *p*), and $(Refx)_k$ *p(x)* for denotation (where $(Refx)_k$ is an operator which turns designation into denotation). $(Refx)_k$ is referred to as the **operator of denotation**, and its linguistic realisations as **applicators**.

Hlavsa distinguishes between the **determination** and the **quantification** component of the denotation operation. He further distinguishes three modes of determination, namely **unique determination** (operator *Un*), **non-unique existential determination** (operator *Ex*) and **non-unique variable determination** (operator *Var*).[20] In the Academic grammar of Czech, correspondingly, the delimitation and quantification attributes of a noun are discerned. The delimitation attribute should characterize the things referred to as specific or non-specific, identical to or different from other mentioned things.

An earlier approach to determination in Czech literature is (Poldauf 1958). It works with features of **specificity** and **identifiability** (for the hearer), like some approaches known from other than Czech literature. Hlavsa relates his three modes of determination to these features in the following way:

- *Un* corresponds to [+specific, +identifiable]
- *Ex* corresponds to [+specific, -identifiable]
- *Var* corresponds to [-specific, -identifiable]

[20] Such a three-way classification of determination had been also proposed in (Bierwish 1970), who distinguishes between semantic features DEFINITE, SPECIFYING and INDEFINITE.

As for the quantificational component, which Hlavsa reflects in the subscript *k* of the denotation operator above, he distinguishes between **definite quantity** (numerical or other), **partial indefinite quantity** (*k:=par*) and **totality** (**k:=tot**).[21]

While some word units used as applicators express just the determination or the quantification component (e.g., in English, the definite article or 'some' for determination, 'all' for quantification), many combine aspects of both in one linguistic expression (e.g., in English, the indefinite article, because it also expresses quantity, or 'both', because it expresses both definiteness and quantity). In Czech, the morphemes *-si, ně-*, which occur for instance in *kdosi, někdo (someone)*, *cosi, něco (something)*, *jakýsi, nějaký (*approx. *some), jaksi, nějak (somehow)*, can be seen as an applicator realizing the *Ex* operator, while *-koli*, which occurs for instance in *kdokoli (anyone), cokoli (anything), jakýkoli (any-arbitrary), jakkoli (anyhow)*, to the *Var* operator.

The differences between the different modes of determination according to Hlavsa are illustrated by the English examples below (Hlavsa 1975, p.14).[22]

(1)

(a) *A hat is an indispensable supplement to lady's* CLOTHING.

(b) *She wants to buy a new* HAT.

(2) *She* SAYS *she bought a* HAT *yesterday.*

(3) *There wasn't* ANYBODY *there, only (a few) hats were laying on the desk and* CHAIRS.

(4)

(a) *Her new hat, which she bought* YESTERDAY, *is very* NICE.

(b) *Joan was nicely* DRESSED, *but that hat did not* SUIT *her.*

In the various occurrences of the noun (group) *hat(s)*, Hlavsa discerns *Var* in (1a,b), *Ex* in (2) and *Un* in (3) and (4a,b).

One difference between Czech and Russian on the one hand, and English on the other hand is that Czech and Russian are languages without a definite and indefinite article. The presence of any applicator is not syntactically obligatory in Czech and Russian in any case.[23] We can therefore speculate, whether a Deictic element should be inserted for every nominal group according to the specification shown above in Figure 21, or whether a Deictic element

---

[21] Hlavsa's view turns out too simplifying. He entirely disregards complex quantifying expressions, e.g., *o málo | o něco méně | více než třikrát tolik* (*little | somewhat less | more than three times as much*), or complex nominal groups like *ani Jan ani Karel* (*neither John nor Charles*), *méně zeleniny než ovoce* (*less vegetables than fruit*), *ne tolik maminek jako babiček* (*not as many mothers as grandmothers*).

[22] As usual, we use the typographical notation of marking the sentence or clause intonation center by SMALL CAPITALS. We make the assumption that the intonation center regularly has the shape of Pierrehumbert's H*L and that its prototypical position is at the last clause part that may be stressed.

[23] According to Hlavsa, a denotation operator is however present in every denotationally used nominal group in Czech, although there may be no overt applicator realizing it. In (Bittner and Hale 1995), it is also argued that such an operator should be assumed in Slavonic languages in general, unlike in Warlpiri.

should be inserted only if its explicit realisation is required by the semantics. One possible solution is to opt for the insertion of a Deictic element in every nominal group, and realise it by an empty string when its explicit realisation is not required. This can be specified as shown in Figure 24.

```
EXPLICIT-DEICTIC
({nominal-specific; nominal-nonspecific}) →
[implicit-deictic] (DEICTIC ! NIL)
[explicit-deictic]
```

Figure 24 Realisation of a deictic element by an empty string

Another possibility is to insert a Deictic element only if it is supposed to be explicitly realised. In this case, we propose to replace the specification shown in Figure 21 and Figure 24 by the specifications shown in Figure 25.

```
NOMINAL
{nominal-relative;                                    nominal-interrogative;
   nominal-specific; nominal-non-specific →
[explicit-deictic] (+Deictic)
[implicit-deictic]
```

Figure 25 Explicit versus implicit deictic element

For instance, when the sentences in (1) through (4) are translated into Czech, as in (5) through (8), none of the occurrences if *klobouk(y)* (*hat(s)*) is accompanied by an applicator, although the same differences in determination as in the English examples are present. The same holds about the Russian translations in (9) through (12).

(5) **Czech**

(a) *Klobouk je nezbytným doplňkem dámského OBLEČENÍ.*

(b) *Chce si koupit nový KLOBOUK.*

(6) *ŘÍKÁ, že si včera koupila nový KLOBOUK.*

(7) *Nikdo tam NEBYL, jenom na stole a na židlích ležely KLOBOUKY.*

(8)

(a) *Nový klobouk, který si koupila VČERA, je velmi pěkný.*

(b) *Jana byla hezky OBLEČENA, jenom klobouk jí NESLUŠEL.*

(9) **Russian**

(a) *Шляпка - необходимая составляющая дамского костюма.*

(b) *Она хочет купить новую шляпку.*

(10) *Она говорит, что вчера купила шляпку.*

(11) *В магазине никого не было, только несколько шляпок лежало на прилавке и на стульях.*

(12)

(a) *Шляпка, которую она купила вчера, очень мила.*

(b) *Джоан была элегантно одета, но шляпка ей не шла.*

As it becomes clear from these examples, a nominal group without any explicit applicator in Czech or in Russian is not by itself specified for determination. However, a correlation can be observed between contextual boundness and specific determination (Hlavsa, Sgall et al.), i.e. *Un*, i.e. contextually bound nominal groups tend to be interpreted as specific (uniquely determining) Therefore, if a contextually bound nominal group is intended as non-specific (non-uniquely determining), such is often made explicit by using an overt deictic element. Non-boundness indicates neither specific nor non-specific determination.[24]

The situation concerning articles is different in Bulgarian. Like Czech and Russian, Bulgarian does not have an indefinite article, but it does have a definite article. The above set of sentences translates as follows into Bulgarian:

(13) **Bulgarian**

(a) *ШапкаТА е необходимо допълнение към женското ОБЛЕКЛО.*

(b) *Тя иска да си купи нова ШАПКА.*

(14) *Тя КАЗА, че вчера си е купила ШАПКА.*

(15) *Там нямаше НИКОЙ, само няколко шапки лежаха на бюрото и столовете.*

(16)

(a) *НоваТА шапка, която тя си купи ВЧЕРА, е много хубава.*

(b) *Жана беше изискано ОБЛЕЧЕНА, но шапкаТА (ТАЗИ шапка) не й ОТИВАШЕ.*

The generalization for Bulgarian seems to be that unique determination is indicated by the definite article, no matter whether the nominal group is contextually bound or not. Nominal groups without an article thus appear non-specific (non-uniquely determining). Bulgarian differs from English in that the definite article is realised by morphological means, namely as a suffix, which is a fact that needs to be taken into consideration in the grammar implementation.

[24] For a definition of contextual (non)-boundness see (Sgall et al. 1986) or the section on thematization and word order in this deliverable.

2.8.2 Determination in the Corpora

We tried to annotate the nominal groups in the Czech AGILE corpus with respect to the Un/Ex/Var distinction. Since most of the nominal groups in our corpus are without any explicit applicator, we had to employ our knowledge of the domain to decide whether a unique object is denoted by a given nominal group. However, it often turned out to be unclear whether unique or non-unique determination was meant, in particular, whether unique or variable determination was at stake. Such ambiguity is quite normal, because in many communicative situations the ambiguity does not matter for the speaker's communicative purpose, as in the following example from the corpus:

(1) *Určete koncový bod dalšího segmentu.*
Lit. *Specify end point next segment*
En. Specify the endpoint of the next segment.

The nominal group *další segment (next segment)* can mean a particular following segment (Un), a non-specified following segment (Ex) or any following segment (Var).

A sidenote: In fact, the text in the original manual is accompanied by a picture which shows three following segments. However, the English original text says *Specify the endpoint of the next segment* (which means Un and nothing else), even though we think it would be more appropriate to say *Specify the endpoint of any following segment* (which would correspond to Var) or *Specify the endpoint of each following segment* (which would correspond Ex). The unique determination used in the English text seems inappropriate, because it seems to entail that any line consists of three segments, or that for every segment there is a segment following it. The Czech wording avoids this entailment, but it would also seem more appropriate to say *Určete koncový bod každého dalšího segmentu. (Specify the endpoint of each following segment)*, thus making explicit that there can be more than one.

It appears that «backward reasoning», trying to guess whether a nominal group was intended to refer uniquely or not, is quite subjective, and therefore we did not annotate the entire corpus in this way and we also do not present the results of the annotation we did. It also appears that recovering the distinctions is not necessary for understanding the texts.

Irrespective of the difficulties we have with identifying the type of determination for some nominal groups in the corpus text, the important observation is that majority of them (at least in Czech and Russian) appear **without any explicit applicator**. The following table summarizes the statistics concerning for the Czech, Russian and Bulgarian corpus.

Number of occurrences of:	Corpus			
	Czech	Russian	Bulgarian	
nominal group (NGR)	667	580	409	
action-denoting nominal group (A)	80	90	19	
object-denoting nominal group (NGR-A)	587	490	390	
object-denoting nominal groups with at least one deictic element	30	16	196	
object-denoting nominal groups with no deictic element	557	464	194	
explicit deictic element (D)	17	10	definite article	194
			Other	13
explicit post-deictic element (PD)	21	14	11	
explicit numerative element (N)	22	14	14	
applicator (D+PD+N)	60	38	232	

Figure 26 Corpus evidence of occurrences of determination and quantification

For instance in the Czech corpus, we found 667 nominal groups (some of them constituting nominal group complexes). Only 30 of them contained some deictic element: according to Halliday's classification, there were 17 occurrences of the correlates of DEICTIC and 21 of DEICTIC_2, and 22 occurrences of NUMERATIVE. So, in total, we encountered 60 applicators in Hlavsa's terms. However, it should be noted that there are is a considerable number of deverbative nominal groups (nominalizations), e.g. *vrácení (returning), uzavření (closing), kreslení (drawing)*, and also other nouns denoting actions, e.g. *návrat (return), dotaz (prompt)*, which do not combine with deictic elements as a matter of principle. Among the 677 nominal groups there are 80 such action denoting ones in our corpus.

Note that in Bulgarian, the action-denoting nominal group (A) as such can have a deictic element —Bulgarian deverbal nouns can take the definite article (or have another deixis in the Hallidayan sense) just as any other noun. This table examines only «**object-denoting nominal groups** with at least one deictic element» or «with no deictic element».

The Bulgarian corpus contains 194 times the definite article and 5 times another deictic, that is we have 199 in total. There are 196 object-denoting nominal groups with at least one deictic element, and 3 deverbal nouns with a definite article.

The examples for Other are the following:

28. Задайте крайната точка на всеки сегмент на полилинията

/ Specify the endpoint of each polyline segment./

49. Можете да въведете тези относителни стойности

/ You can enter these relative values/

84. изберете линия или дъга за всеки край на мултилинията

/ select a line or an arc for each end/

93. този цвят не се показва …

/ this color does not show/

122. За да видите списък на всички стилове …

/ to list all styles/
един от следните методи (3,22,36,58,113,143,157,173)
one of these methods

Now we would like to combine the insights offered by Hlavsa for Czech with Halliday's treatment proposed for English, which is also incorporated in the KPML grammar, and extend this integrated view also for Russian and Bulgarian. First of all, there are obvious differences in terminology, so we shall try to project Hlavsa's ideas into Halliday's system.

Hlavsa distinguishes between determination and quantification not only when they are expressed by separate words (expressions) but also when they are reflected by morphemes within one word. Halliday in fact does this too, because he distinguishes between singular and non-singular deictic items.

Halliday distinguishes between specific and non-specific deictic items. Hlavsa has a three-way classification of determination into unique determination (operator *Un*), non-unique existential determination (operator *Ex*) and non-unique variable determination (operator *Var*). At first sight, it appears that specific determination should (approximately) correspond to unique determination, and non-specific determination to non-unique (existential or variable) determination. Let us now attempt a more detailed comparison, in order to be able to place the Slavonic applicators into Halliday's classification and thus subsequently into the KPML system(s).

2.8.3 Specific Determination

The notion of specific determination in Halliday's terms appears to mean the same as **unique determination** in Hlavsa's terms, because it reflects that a **particular subset** (class of things) is under question and that it is **identifiable** both for the speaker and the hearer.

2.8.3.1 Specific Deictic Items in Czech

Halliday classifies also the possessive deictics as specific deictic items. However, in Czech they can appear in combination with deictic items, both specfic or non-specific, e.g. *tato moje kniha* (lit. *this my book,* En. *this book of mine*), *tato Janina kniha* (lit. *this Jane's book,* En. *this book of Jane's*), *nějaká moje kniha* (lit. *some my book,* En. *some book of mine*), *všechny moje knihy* (*all my books*), *každá moje kniha* (lit. *every my book,* En. *every book of mine*), *jakákoli Janina kniha* (lit. *any Jane's book,* En. *any book of Jane's*). Also, nominal groups like *moje kniha* or *Janina kniha* cannot be considered to denote uniquely.

In our corpus, we encounter one occurrence of a possessive pronoun in segment 177: *pro její ukončení* (Lit. *for its ending*, En. *to end it*), where the pronoun co-refers to a multiline. In this case, the entire nominal group can be considered to refer to a particular action (Un), but this is not because of the possessive pronoun per se, but because the drawing of an object can only be ended once (unless we cosider repetitive actions).

Hlavsa does not discuss possessives among applicators, and also the Academic grammar does not list possessives among either specific or non-specific attributes of delimitation. One interpretation of this fact is that they do not consider them to belong into this class. And this is the stance we shall adopt, so we shall include possessives under neither specific nor non-

specific deictic items. Rather, we shall assume a distinct function Possessive in the experiential structure of a nominal group in Czech.[25]

To make this discussion complete, we shall also consider interrogative items, even though they are not really relevant in the context of our project, because we are not generating questions. It is interesting to note that while Halliday includes *which(ever), what(ever)* and *whose(ever)* under specific deictics, Hlavsa discusses the Czech equivalents of the first two, i.e. *který* and *jaký*, under the heading of non-unique existential determination. Hlavsa's view, based on Gebauer's and Ertl's, is substantiated by the denotational equivalence between indicative and interrogative sentences as in (a) and (b) in the following examples:

(2)

(a) *Který podnik vám opravoval výtah?*
(Which company repaired your elevator?)

(b) *Některý|nějaký podnik vám opravoval výtah.*
(Some company repaired your elevator.)

(3)

(a) *Jaký|který utopistický román Petr četl?*
(What|which utopic novel was read by Peter?)

(b) *Petr četl nějaký|některý utopistický román.*
(Peter read some utopic novel.)

If we accept this argument based on denotational equivalence, we should consider the interrogatives as non-specific deictic items in Halliday's classification.

The Czech interrogative possessives, e.g. *čí* (*whose*), again are not discussed by Hlavsa or in the Academic grammar. We propose to treat them in the same way as the determinatice possessives, i.e. by assigning a separate function Possessive to them in the experiential structure.

It should be noted that interrogatives seem incompatible with deictics, whether specific or non-specific: a nominal groups like *čí tato kniha* (*whose this book*) or *čí některá|nějaká kniha* (*whose some book*) is not acceptable. Therefore, if we consider interroggative possessives separately from deictics, we have to ensure that the grammar encodes their mutual exclusivity.

On the basis of the above discussion, we conclude that the repertoire of specific deictic items in Czech is narrower than that for English. We summarize the specific deictic items in Czech in the table below, which is adapted from (Halliday 1994, p. 181).[26]

---

[25] This view will be further supported by considering the principles of ordering within nominal group in Czech, where we shall see that we cannot include Possessives with other Epithets or Classifiers either.

[26] Pronouns, demonstrative as well as possessive, in Czech fall into the grammatical category of adjectives, which in turn belong to the general class of nouns. As such, their forms are distinguished for Gender, Number and Case. The same holds for all other adjectives and nouns. When we list deictic and other items in tables, we include the masculine forms in nominative case; when number is not distinguished, we list the singular form(s). See more on agreement within a nominal group below.

Specific deictic items (Czech)		***Determinative*** singular/mass	dual/plural
Demonstrative	proximity-unmarked	*ten, onen*	*ti*
	proximity-near	*tento, tenhle, tenhleten*	*tito, tyhleti*
	proximity-far	*tamten, tamhleten*	*tamti, tamhleti*

Figure 27 Items functioning as specific Deictic in Czech

Syntactically, all these items are demonstrative pronouns which agree in grammatical number, gender and case with the Thing.

2.8.3.1.1 Evidence from the Czech Corpus

The occurrences of specific deictics in our corpus are enumerated in the next table below. The deictic elements are underlined. The numbers in brackets denote the number of segment in our corpus where the nominal group occurs.

Nominal group with specific deictic		***Anaphoric antecedent***	
jedním z těchto způsobů (2)	*using one of these methods*	--	--
tyto relativní hodnoty (65)	*these relative values*	*vzdálenost a úhel* (64)	*distance and angle*
v tomto případě (66)	*in this case*	--	--
tento rozměr (87)	*this offset*	*rozměr posunutí* (86)	*the offset*
typ čáry daného elementu (96)	*the element's linetype*	*element* (91)	*an element*
tato barva (126)	*this colour*	*barva pozadí* (125)	*a background colour*
nastavení tohoto stylu (138)	*(to) set this style*	*vytvořený styl multičáry* (137)	*the newly created mline style*
ty (čáry), které již byly nakresleny (231)	*those (lines) already drawn*	--	--

Figure 28 Occurrences of specific Deictic items in the Czech corpus

It is apparent that the nominal groups in which specific deictics appear are used for textual co-reference. Most of them are used to refer back in the text (anaphorically), and they have explicit antecedents which we also include in the table. The specific deictics encountered in the corpus are all various forms of *tento* (*this*), except for the one case where the *daný* (lit. *given*) is used in *typ čáry daného elementu* (*type of line of the element*). The function of the nominal group *daný element* is to co-refer. The antecedent in this case realises the variable mode of determination.

One exception is the nominal group *v tomto případě* (*in this case*) is not easy to characterize as for its referential status. It appears within a sequence of instructions of an example. Most likely it is meant to refer to the situation in the context of the example, the picture being drawn as depicted in the accompanying figure. So we can either consider it to refer anaphorically to the example as such, or we can say it does not have an explicit antecedent in the text, in which case it would constitute exophoric reference.

Another exception is the nominal group *jedním z těchto způsobů* (*one*-instr *of these*-gen *ways*-gen) where *tyto způsoby* (*these ways*) refers forward (cataphorically) to the following passage of the text. Note that the nominal group *jeden z těchto způsobů (one of these way)* as such does not refer uniquely, it expresses existential determination according to Hlavsa. A similar nominal group with the same function occurs a number of time throughout the corpus, but *následující* (*following*) is used instead of *tyto* (*these*), e.g., *jedním z následujících způsobů* (*one*-instr *of following*-gen *ways*-gen) in 28, 52, 148, 180, 195, 212, and jednou z následujících metod (*one*-instr *of following*-gen *methods*-gen) in 79. The adjective *following* is not included in Halliday's classification of deictic elements, but it seems reasonable to consider it to belong to the DEICTIC_2 group.

Last but not least, we also included the complex nominal group *ty, které již byly nakresleny* here. It occurs in the context of *čára, kterou právě kreslíte a ty, které již byly nakresleny* (*the line you're drawing and those already drawn*). In agreement with Halliday, we consider it a case of ellipsis where the deictic element *ty* has been «promoted» to serve as the Head of the nominal group, since the Thing which corresponds to čáry (lines) has been elided.

So it seems that an explicit specific deictic element is used in the Czech texts in our corpus only in cases where it is necessary to explicitly mark a «phoric» nominal group (anaphoric, cataphoric or exophoric). This is not so for every «phoric», esp. anaphoric nominal group, so we find many of them without a deictic item too. The contextual conditions under which a deictic is necessary should be a topic for further research within the Work Package 5 (Text Structuring).

2.8.3.1.2 Formalization of Specific Deictics in Czech

We specified earlier how a Deictic element gets inserted and possibly realised by an empty string (see Figure 21 and Figure 24, or Figure 25). Now we specify how to arrive at a particular explicit realisation of a specific deictic element in Czech, depending on the proximity feature. On the basis of the discussion provided above, and referring to the summary table of Czech specific deictic items in Figure 27, we propose the initial formalization in terms of systemic networks shown in Figure 29. The choice between various realisations of the demonstrative deictic elements needs to be textually determined. The default choices are listed first.

```
SPECIFIC-TYPE
(nominal-specific, explicit-deictic)→
     [demonstrative-selection]
```

```
DEMONSTRATIVE-NONQUESTIONING
({demonstrative-specific-pronoun; demonstrative-selection})→
     [proximity-neutral]
     [proximity-near]
     [proximity-far]

TEN-ONEN-DANY
(proximity-neutral) →
     [TEN] Deictic ! TEN
     [DANY] Deictic ! DANY
     [ONEN] Deictic ! ONEN

TENTO-TENHLE-TENHLETEN
(proximity-near) →
     [TENTO] Deictic ! TENTO
     [TENHLE] Deictic ! TENHLE
     [TENHLETEN] Deictic ! TENHLETEN

TAMTEN-TAMHLETEN
(proximity-far) →
     [TAMTEN] Deictic ! TAMTEN
     [TAMHLETEN] Deictic ! TAMHLETEN
```

Figure 29 Formal specification of Czech specific determination

2.8.3.2 Specific Deictic Items in Russian

The situation is Russian is quite similar to that in Czech. The following table provides a summary of the specific deictics in Russian.

<table>
<tr><td colspan="3" rowspan="3">Specific deictic items</td><td colspan="3">Determinative</td><td rowspan="3">Interrogative</td></tr>
<tr><td colspan="3">Proximity</td></tr>
<tr><td>near</td><td>far</td><td>neutral</td></tr>
<tr><td rowspan="3">Demonstrative</td><td>selective</td><td>singular/mass</td><td rowspan="2">этот</td><td rowspan="2">тот</td><td>данный</td><td>который</td></tr>
<tr><td></td><td>plural</td><td>такой</td><td></td></tr>
<tr><td colspan="2">non-selective
non-pronominal</td><td></td><td></td><td>(word order)</td><td>какой</td></tr>
<tr><td rowspan="2">Possessive</td><td>pronominal</td><td>person</td><td colspan="3">мой, твой, наш, ваш, его, ее, их, свой
чей-то</td><td>чей</td></tr>
<tr><td colspan="2">non-pronominal</td><td colspan="3">possessive adjectives</td><td></td></tr>
</table>

Figure 30 Items functioning as specific Deictic in Russian

2.8.3.2.1 Evidence from the Russian Corpus

According to the Table, the determiners as items are rare in Russian. There are only three occurrences of specific ones, one possessive, quite normal, and some cases with *следующий*

(next).In Russian the determination is close to the word order and interrelated with some other categories (aspect, for example). The following table overviews the occurrences of specific deictic items in the Russian AGILE corpus.

Nominal group with specific deictic		***Anaphoric antecedent***	
одним из следующих способов (3)	*using one of these methods*	--	--
в следующей форме (58)	*in the form*	--	--
эти данные (58)	*these relative values*	*расстояние и угол линии*	*distance and angle*
этот элемент (71)	*the element*	*смещение элемента линии [69].*	*the offset of the line element*
чтобы дуга проходила через ее конечную точку(45.9).	*so the arcs snaps to the endpoint of the line.*	*прямая*	*line*
укажите этот стиль (103)	*(to) set this style*	*вновь созданный стиль мультилинии (102)*	*the newly created mline style*

Figure 31 Occurrences of specific (and possessive) Deictic items in the Russian corpus

2.8.3.2.2 Formalization of Specific Deictics in Russian

The situation is very similar to Czech, though the lexical choice is much more restricted. Selection of [context-bound] should lead to word-order alternations, for example, driving this element to the left.

```
SPECIFIC-TYPE
(nominal-specific, explicit-deictic)→
     [demonstrative-selection]

DEMONSTRATIVE-NONQUESTIONING
({demonstrative-specific-pronoun; demonstrative-selection})→
     [proximity-neutral]
     [proximity-near]
     [proximity-far]

PROXIMITY-NEUTRAL-TYPE
(proximity-neutral) →
     [DANNYJ] Deictic ! DANNYJ
     [TAKOJ] Deictic ! TAKOJ
     [context-bound] Deictic ! ELLISPSISZERO

TENTO-TENHLE-TENHLETEN
(proximity-near) →
     [ETOT] Deictic ! ETOT

TAMTEN-TAMHLETEN
(proximity-far) →
     [TOT] Deictic ! TOT
```

Figure 32 Formal specification of Russian specific determination

2.8.3.3 Specific Deictic Items in Bulgarian

Using the definite article in Bulgarian is the equivalent of the English *the* as illustrated in the first line of the table to follow. Its realisation is by means of a postpositional noun affix marked for gender and number. The first component of the noun phrase which is non-specific deictic receives the marker/affix.

It is possible to have phrases such as *ТАЗИ червенаТА линия* in some contexts in Bulgarian, for example, when identifying one of several possible referents of near proximity. Such cases are expressed by a combination of selective (*ТАЗИ*) and non-selective determinative demonstrative deictic item (*червенаТА*) in the same nominal group.

It appears that Possessives can be treated separately from deictic items as proposed above in the discussion of Czech deictics, since they may form similar constructions as those with adjectives, as in Czech. It is also possible to combine a Possessive, a demonstrative deictic and the definite article, e.g., *ТАЗИ моятА червена линия* (*THIS THE mine red line*).

Specific deictic items (Czech)		***Determinative*** singular/mass	dual/plural
Demonstrative	proximity-unmarked	*-ът, -ят,-а, -я -та, -то*	*-те*
	proximity-near	*този, (моя), тази,(тая) това*	*Тези, тия*
	proximity-far	*онзи, (оня), онази, (оная), онова*	*Онези, ония*

Figure 33 Items functioning as specific Deictic in Bulgarian

2.8.3.3.1 Evidence from the Bulgarian Corpus

In the studied corpus there are no examples of combining selective and non-selective determinative demonstrative deictic items in the same nominal group. The following table contains corpus occurrences of specific deictic items without articles and possessive deictics.

Nominal group with specific deictic		***Anaphoric antecedent***	
Тези относителни стойности (49)	*these relative values*	*разстоянието и ъгъла* (48)	*distance and angle*
Този цвят (93)	*this colour*	*основния цвят* (91)	*a background colour*

Figure 34 Occurrences of specific Deictic items in the Bulgarian corpus

In the Bulgarian corpus, many of the examples which in English are classified under ***specific deictics***, especially ***Demonstrative*** proximity-near, in translation appear as the so-called postdeictics (Deictic2), for example:

using one of these methods	*като използвате един от следните методи (3)*
	/using one of the following methods/
in this case	*в дадения случай (49)*
	/in the given case/

2.8.3.3.2 Formalization of Specific Deictics in Bulgarian

```
NOMINAL
{nominal-relative; nominal-interrogative;
      nominal-specific; nominal-non-specific →
[explicit-deictic] (+Deictic)
[implicit-deictic]

SPECIFIC-TYPE
(nominal-specific, explicit-deictic)→
     [nonselective] Deictic ! Definite-article
     [demonstrative-selection]

DEMONSTRATIVE-NONQUESTIONING
({demonstrative-specific-pronoun; demonstrative-selection})→
     [proximity-near]
     [proximity-far]
```

Figure 35 Formal specification of Bulgarian specific determination

2.8.4 Partial Non-Specific Determination

Halliday's class of non-specific deictics appears to correspond to Hlavsa's class of applicators realizing non-unique determination, both existential and variable. Among non-specific deictic items, Halliday distinguishes further, though without any discussion, between **total** and **partial**, and among the latter he makes a further distinction between the **selective** and **non-selective** ones. We can ask whether there are any correspondences between Halliday's and Hlavsa's subclasses.

One difference is that Hlavsa considers totality and partiality as distinctions concerning the quantification component of delimitation, orthogonal to determination. The operation of quantification «curves out» a quantity of entities from the domain of the designator or from some given set. Two oppositions play a role in characterizing the quantity of entities, namely the opposition **definite : indefinite** number and **total : partial** quantity.

Definite quantity is normally expressed by numerals, which are separate applicators. The corresponding function in Halliday's system is the NUMERATIVE.

The situation is the same in Czech, Russian and Bulgarian as in English, that a deictic item does not express a definite quantity; an exception is the deictic item corresponding to *both* in all the three languages (Czech: *oba*, Russian: *оба*, **Bulgarian:** *оба*) which expresses the definite quantity of two.

Indefinite quantity, on the other hand, can be expressed by a variety of means; probably most of the applicators discussed by Hlavsa would belong to NUMERATIVE in Halliday's system, but a few are mentioned as Deictic items.

Hlavsa also discusses certain very tight relations among determination on the one hand and indefinite (total versus partial) quantification on the other hand. So it seems possible to follow Halliday's approach and merge the applicators of determination and indefinite quantification into one classification scheme, as we shall do in this section.

Adopting the distinction between existential and variable mode of non-specific determination can be reflected in a formal specification as shown in Figure 36. At the same time, we need to maintain the distinction between partial and total deictic items, as shown in Figure 39.

```
NONSPECIFIC-TYPE-I
{NOMINAL-NONSPECIFIC; NONDEICTIC-VERBAL-AGENT} →
   [existential]
   [variable]
```

Figure 36 Formalization of the distinction between existential and variable non-specific determination

```
NONSPECIFIC-TYPE-II
{NOMINAL-NONSPECIFIC; NONDEICTIC-VERBAL-AGENT} →
   [partial]
   [total]
```

Figure 37 Formalization of the distinction between partial and total non-specific determination

Let us first overview Hlavsa's characterization of applicators realizing existential and variable non-unique determination. They fit into Halliday's system as partial non-specific Deictic items. Then we will discuss Hlavsa's applicators realizing indefinite total quantification, which correspond to total non-specific Deictic items in Halliday's scheme.

2.8.4.1 Partial Existential determination in Czech

With this mode of determination, as described in (Hlavsa 1975), the identifiability of the object is not expressed, although the conditions for a unique identification of the object have been satisfied, and it is possible that such happens later in the communication.

The primary applicators realizing existential non-unique determination in Czech are *nějaký, jakýsi* (*some*) and unstressed *jeden* (*one*). The unmarked one among them is *nějaký*, which unlike the others can also realise variable determination. The difference in the distribution of *jakýsi* and *jeden* is described by Hlavsa, and also by Poldauf (1969), using a feature of **familiarity** which has a positive value for *jeden* and a negative value for *jakýsi*. Moreover, *jeden* can only combine with singular nouns. Applicators with similar function as *jeden* are *jistý* (*certain*) and *určitý* (*certain, specific*). The following examples are offered by Hlavsa:

(4) *Chtěl by si někdo dopisovat s jedním Američanem?*
(Would someone like to be penfriends with one American?

(5)

(a) *Jan ještě zajde pro jednoho přítele. (John will still go there for one friend.)*

(b) *Jan ještě zajde pro nějaké přátele. (John will still go there for some friends.)*

(6) *To léčení mu poradila jakási kořenářka.*
(The treatment was recommended to him by some medicine-woman.)

In (4) and (5), the speaker uttering *jeden* would be able to identify, or further describe, the object; in (5), it is also possible that the speaker takes over the perspective of another

person, i.e the identity of 'jeden přítel' (one friend) is known to Jan, not to the speaker. It is more usual with *jeden* than with other existential applicators that the identity of the object is specified in the following context. Concerning *jakýsi*, it should also be remarked that it has an additional optional meaning of «insufficient», «not good enough», as in (6).

Note that *nějaký* and *jakýsi* are created from the same base-form, namely *jaký*, using the prefix *ně-* or the suffix *–si*, respectively. Another pair of non-unique existential applicators can be created from the base-form *který*, namely *některý* and *kterýsi*. Using the suffix *–koli*, two variable applicators can be created from these two bases, namely *jakýkoli* and *kterýkoli*. Both *jaký* and *který* mean roughly *some* in English; more precisely, *jaký* means 'having what property', 'being some kind of', whereas *který* means more 'which one'. Křížková (1971, p. 163-4) states that the general meaning of the interrogative pronoun *jaký* in all slavonic languages reflects a static feature, whereas *který* reflects individualization. When we stretch this to the applicators, we can say that *nějaký*, *jakýsi* and *jakýkoli* express that further features of the named object are unspecific or unknown, whereas *některý*, *kterýsi* and *kterýkoli* express that individual characteristics of the object are unspecific. This difference is apparent from the way such uncertainty or lack of specificity can be removed. With nominal groups containing the former applicators, created from *jaký*, further attributes, e.g. *nějaké šaty* (*some kind of dress*)→ *letní šaty* (*summer dress*), hyponyms, e.g. *nějaký strom* (*some kind of tree*) → *jabloň* (*apple tree*), or unique denotation can do this job. With nominal groups containing the latter applicators, created from *který*, only unique denotation can remove the uncertainty, e.g. *některá kniha* (*some book*) → *ta kniha* (*the book*) or *Krakatit*.

Hlavsa thus introduces a feature called «knowledge of an individual characteristic». The applicators based on *jaký* are unmarked with respect to this feature, whereas the ones based on *který* are marked. The following paralel subsystems can be discerned in Czech:

Knowledge of individual characteristic	**Applicators for Ex**	**Applicators for Var**
unmarked	*jakýsi, nějaký*	*jakýkoli, nějaký*
marked lack thereof	*kterýsi, některý*	*kterýkoli, některý*

Figure 38 Classification of Czech deictics according to knowledge of individual characterictic

According to (Hlavsa 1975, p. 63), marked lack of knowledge of an individual characteristic presumes that the scope of possible denotation be ostensively, relationally or contextually established. Therefore, «*některý|kterýkoli x*» actually means «*některý|kterýkoli z Un množiny x*» (*some|any from Un set x*).

Especially with regard to the AGILE project, it is important to note that Hlavsa quotes some authors, e.g. (Křížková 1971), who noticed that the applicators containing the morphological component *-si*, e.g. *jakýsi, kterýsi* (*some*), but also the pronouns *kdesi* (*somewhere*), *kdosi* (*someone*), etc., can only be used under specific syntactic conditions. Hlavsa proposes his generalization that existential applicators do not appear in interrogative and imperative sentences. To support his claim, he points out that he following sentences are ill-formed:[27]

[27] When one uses nějaký, někde, někdo in (11,12,13), their mode of determination can only be variable.

(7)

(a) * *Koupíte si jakousi knihu? (Will you buy one book?)*

(b) * *Kupte si jakousi knihu! (Buy one book!)*

(8)

(a) * *Je kdosi doma? (Is someone at home?)*

(b) * *Ať je kdosi doma! (Let someone be at home!)*

(9)

(a) * *Počkáš na mne kdesi? (Will you wait for me somewhere?)*

(b) * *Počkej na mne kdesi! (Wait for me somewhere!)*

The reason for this might be that in such contexts the speaker has to clearly specify either which (identifiable) object she has in mind or whether an arbitrary object is concerned. In any case, we should exclude the use of existential applicators in imperative mode in the texts generated in AGILE.

We would like to be able to place the non-unique existential applicators mentioned so far into Halliday's scheme which distinguishes between selective and non-selective non-specific deictic items. We have the following two ideas:

A. The question is whether Hlavsa's **knowledge of individual characteristics** can be straightforwardly mapped onto Halliday's **selectivity**. If so, it seems that «marked lack of individual characteristics knowledge» would correspond to non-selectivity. So, *kterýsi, některý* (*some*), *kterýkoli* (*any*) would be classified as non-selective, while *jakýsi* (*some*), unstressed *jeden* (*one*), *jistý* (*certain*), *určitý* (*certain, specific*) and *jakýkoli* (*any*) would be classified as selective. This seems natural.

B. Another possibility is that the distinction between selective versus non-selective dectics according to Halliday would correspond to the distinction between existential versus variable determination according to Hlavsa, so all the abovementioned applicators would belong to the class of selective ones. The variable applicators would belong to non-selective deictics. This also makes sense as for the interpretation of the term «selectivity».

However, in Halliday's system, it is the non-selective (and non-singular) deictics that combine with mass nouns. But the Czech *kterýsi, některý* which we would classify as non-selective in approach A do not. Under approach B we do not get a simple mapping either. Actually, the best description of the facts probably is that the applicators based on *který* do not combine with mass nouns, while the ones based on *jaký* do. It appears that *určitý* also does, while *jistý* and *jeden* do not. If we specify it in this way, approach A seems preferable over approach B. These claims are summarized in the table below

Partial existential deictic items (Czech)		unmarked	non-mass
selective (unmarked for knowledge of individual characteristic)	familiarity	--	*jeden* (*one*) *jistý* (*certain*)
	non-familiarity	*jakýsi, nějaký* (*some*) *určitý* (*certain, specific)*	
non-selective (marked for lack of knowledge of individual characteristic)		--	*kterýsi, některý* (some)

Figure 39 Partial Non-Specific Existential Deictic items in Czech

2.8.4.1.1 Evidence from the Czech Corpus

The Czech corpus does not contain any occurrences of *jakýsi* (*some*) which is not surprising given the preceding claim. Our corpus also does not contain any occurrences of *nějaký* (*some*), although it would be possible to use them at some places without any change in the meaning.

As noted above, we do find nominal groups containing unstressed *jeden* (*one*) expressing non-unique determination in our corpus. There also appear several nominal groups containing *určitý* (*certain*). We summarize the occurrences of the non-unique existential applicators in the table below.

Nominal groups with an existential applicator	
jedním z těchto způsobů (2)	*using one-instr of these-gen ways-gen*
jedním z následujících způsobů (28, 52, 148, 180, 195, 212)	*using one-instr of following-gen ways-gen*
jednou z následujících metod (79)	*using one-instr of following-gen methods-gen*
kopie úseček v určité vzdálenosti na určitou stranu nebo procházející určitým bodem (25)	*copies of lines offset at a specified distance to one side or through a point*
nový objekt v určité ekvidistantní vzdálenosti od vybraného objektu nebo objekt procházející určitým bodem (146)	*new objects at a specified offset from a selected object or through a specified point*
jiný rovný úsek (46)	*another straight segment*
Každý element může mít různou barvu a typ čáry. (105)	*Every element can have a different color and linetype.*

Figure 40 Occurrences of nominal groups with an explicit existential determiner in the Czech corpus

The English texts are the texts from the English manual. We have underlined the corresponding deictic elements in the Czech and English versions. Note that the Czech *určitý* is used in the translation of *one side* and several times of *a specified*. As for the latter,

the translation of in the Czech manual is not literal, because *specified* translates literally as *určený* or *daný* and not *určitý*. On the other hand, the chosen wording does capture the existential determination, namely that the identity of the object(s) will be known to the user. Alternative wordings with *nějaký* or *jistý* would also be acceptable. The second occurrence of *určitý* in segment 25 could also be replaced by *jeden*. All the occurrences of *určitý* could be replaced by *(za)daný* (*specified*) preserving the same meaning, while if *určený* (*specified*) were used, the sentences would sound awkward; better sounding wording would then be obtained with *nějaký určený* (*some specified*). This perhaps supports the view that *daný* can function as DEICTIC$_2$, but *určený* not, it functions just as an EPITHET.

Applicators based on *který*, i.e. *některý, kterýkoli,* do not appear in our corpus, even though the functional conditions for their use are sometimes fullfilled. For instance, the nominal groups *jedním z těchto|následujících způsobů* could be replaced by *některým z těchto|následujících způsobů.* However, *jeden* is prefered over *některý* in singular nominal groups as noted by Hlavsa.

Thus, *některý* would be really appropriately used with plural nominal groups or when a repetitive situation is denoted, for example in the made-up sentence *Některý údaj | Některé údaje se mění při každém ukládání* (*Some data change with every saving*).

The applicators *jiný* and *různý* in the last two lines in the table realise comparative reference, and would be treated as POST-DEICTIC in Halliday's system. We leave them out in the current specification, because we have not included comparative reference yet.

2.8.4.1.2 Formalization of Partial Existential Deictics in Czech

In the following table, we propose a formalization of the partial existential determination in Czech on the basis of the above discussion.

```
EXISTENTIAL-PARTIAL-SELECTIVE-TYPE
(existential, partial, selective) → [familiar]   [non-familiar]

JEDEN-JISTY
(familiar) →
   [JEDEN] (Deictic ! JEDEN)
   [JISTY] (Deictic ! JISTY)

JAKYSI-URCITY
(non-familiar) →
   [JAKYSI] (Deictic ! JAKYSI)
   [URCITY] (Deictic ! URCITY)

NEJAKY
(partial, {non-familiar; variable}) → [NEJAKY] (Deictic ! NEJAKY)

EXISTENTIAL-PARTIAL-NONSELECTIVE-TYPE
(non-selective, existential) → [KTERYSI] (Deictic ! KTERYSI)

NEKTERY (partial, non-selective) → [NEKTERY] (Deictic ! NEKTERY)
```

Figure 41 Formalization of existential partial non-specific determination

2.8.4.2 Partial Existential determination in Russian

Existential non-unique determination in Russian is quite similar to Czech. Basically, in Russian this is expressed by *какой-то* (*some*) and unstressed *один* (*one*). Similarly to Czech, the difference in the distribution of *какой-то* and *один* depends on a feature of

familiarity which has a positive value for *один* and a negative value for *какой-то*. Moreover, *один* can mostly specify singular nouns. Deictics with a similar function as *один* are *некоторый* (*some certain*) and *определенный* (*certain, specific*).

While Czech has the parallel systems of applicators based on *jakýˊ* and *kteryˊ* which are distinguished by the knowledge of an individual characteristic (selectivity), Russian does not, since it does not have the equivalent of *jakýˊ*. The Russian existential applicators are summarized in the table below.

Partial existential deictic items (Russian)	non-mass
familiarity	*один, некоторый*
non-familiarity	*некий, какой-то*

Figure 42 Partial Non-Specific Existential Deictic items in Russian

2.8.4.2.1 Evidence from the Russian Corpus

Nominal groups with an existential applicator	
одним из следующих способов (3, ...)	*using one*-instr *of following*-gen *methods*-gen
на указанное расстояние в одну сторону от точки или через точку (24)	*at a specified distance to one side or through a point*

Figure 43 Occurrences of nominal groups with an explicit existential determiner in the Russian corpus

2.8.4.2.2 Formalization of Partial Existential Deictics in Russian

In the following table, we propose a formalization of the partial existential determination in Russian on the basis of the above discussion.

```
EXISTENTIAL-PARTIAL-SELECTIVE-TYPE
(existential, partial, selective) → [familiar]   [non-familiar]

FAMILIAR-TYPE
(familiar) →
   [ODIN] (Deictic ! ODIN)
   [NEKOTORYJ] (Deictic ! NEKOTORYJ)

NON-FAMILIAR-TYPE
(non-familiar) →
   [KAKOJ-TO] (Deictic ! KAKOJ-TO)

NEKIJ
(partial, {non-familiar; variable}) → [NEKIJ] (Deictic ! NEKIJ)

EXISTENTIAL-PARTIAL-NONSELECTIVE-TYPE
(non-selective, existential) → [KOTORYJ] (Deictic ! KOTORYJ)

KAKOJ-NIBUDJ
(partial, non-selective) → [KAKOJ-NIBUDJ] (Deictic ! KAKOJ-NIBUDJ)
```

Figure 44 Formalization of existential partial non-specific determination

2.8.4.3 Partial Non-Specific determination in Bulgarian

In Bulgarian language there are no different deictic items for Existential and Variable partial non-specific determination.

The main system for non-specific deictics could be:

NONSPECIFIC-TYPE
{NOMINAL-NONSPECIFIC} →
[partial]
[total]

Non-Specific deictic items		**Number**		
		Singular	plural	
			mass/plural	unmarked
Partial	Selective	*Един (one)* *Някой (either)*	*Едни (some)*	*Някои (some,any)*
	non-selective	---	---	

Figure 45. Partial Non-Specific Deictics in Bulgarian

In general, the lack of definite article (marked morphologicaly by suffics when presented) in the nominal groups is the sign of «Non-specifity» in Halledeyan terms and carries the same meaning as English non-selective deictic a(n) for nominal groups in singular.

2.8.4.3.1 Evidence from the Bulgarian Corpus

1) Occurrences of nominal groups in singular «non-selected»

- titles of the procedures (1, 20, 33, 56…) :
Чертаене на линия *To draw a line*

-most frequently in the indicative clauses of the corpus:
Можете да започнете нова линия (17) *You can start a new line*
ако искате да се появи линия (83) *to display a line*
За да изберете стил (118) *to select a style*
…
-there are some occurences in the imperative clauses:
Изберете елемент (68) *Select an element*
посочете основен цвят (95) *select the background fill color*

2) Non-selective plural
самостоятелни сегменти от линии и дъги (55) *individual line and arc segments*
повторете стъпки 4-7 (78) *repeat steps 4 though 7*
…

3) Selective singular
един от следните методи (3,22,36,58,113,143,157,173) *one of these methods*

2.8.4.3.2 Formalization of Partial Non-Specific Deictics in Bulgarian

```
PARTIAL-PLURALITY-TYPE
(partial) →
     [singular]
     [plural]

PARTIAL-SELECTIVE-TYPE
(partial) →
     [selective]
     [non-selective] (Deictic ! NIL)

PARTIAL-SELECTIVE-SINGULAR
(singular, selective) →
     [EDIN] (Deictic ! EDIN)
     [NIAKOI-S] (Deictic ! NIAKOI-S)

PARTIAL-SELECTIVE-PLURAL
(plural, selective) →
     [NIAKOI-P] (Deictic ! NIAKOI-P)
```

Figure 46. Formalization of the partial non-specific determination in Bulgarian

2.8.4.4 Partial Variable determination in Czech

We have already mentioned variable denotation in the discussion of applicators based on *jaký* versus *který*. Let us now complete the overview of Hlavsa's approach to determination by discussing this last more in more detail.

With this mode of determination, as described in (Hlavsa 1975), the identity of the denoted element or set of elements is not specified, the choice is arbitrary within the given domain. Therefore, the variable determination cannot occur in clauses which as such express unique determination, i.e. are bound to a particular uniquely placed situation. Cf. the following Hlavsa's examples:

(10) *Tatínek jezdí kteroukoli tramvají. (Father takes any tram.)*

(11) * *Tatínek právě jede domů kteroukoli tramvají.*
(Father is now taking any tram home.)

The difference between (10) and (11) is that the former talks about any of father's tram-trips, as if expressing his habits. The latter is supposed to describe the individual tram-trip

which is taking place at the time of the utterance, and therefore the variable determination is not appropriate. It would be appropriate to use existential determination.

The characterisic of clauses as for whether they are bound to a particular situation or not is referred to as «actuality» (Cz. **aktuálnost, neaktuálnost**) by Hlavsa. In Czech, the clause (i.e. the verb) does not express «actuality» obligatorily. Therefore, a variable applicator used for naming and object can in fact signal «non-actuality», but the opposite does not hold: «non-actuality» determined by other means does not necessarily allow for using a variable applicator, e.g. *kterýkoli* or *jakýkoli*. This can be illustrated with the following examples, the first of which is ambiguous as for actuality, but the second is not, it can only be interpreted as non-actual:

(12) *Nakupovala v tomto obchodě.*
She used to shop in this store or She shopped in this store.

(13) *Nakupovala v kterémkoli obchodě.*
She used to shope in any store.

The primary means of expressing variable determination are the applicators created with the suffix *-koli*, e.g. *jakýkoli* and *kterýkoli*. It appears that also the adjective *libovolný* (*arbitrary*) can fulfil this function. In comparison to the other modes of determination, the means used to express variable determination are the most varied among the three modes of determination, the least uniform in Czech. Hlavsa states that the English *any* has precisely the function of variable determination and is much more uniformly used.

When trying to annotate nominal groups without any explicit applicator in a corpus, it is often a matter of analysis (and reasoning) to find out whether variable determination is at stake. There are some negative signals which can be used in analysis, and appear useful in our AGILE context. One is that variable determination is only compatible with verb-predicates which carry the feature of possible repetition. Therefore, for instance, (14) is fine, but (15) which only describes one action is ill-formed:

(14) *Příkaz ČÁRA nakreslí (jakoukoli) křivku. (Command POLYLINE draws (any) line.)*

(15) * *Příkaz ČÁRA nakreslil jakoukoli křivku. (Command POLYLINE drew any line.)*

Another restriction is that is the verb-predicate expresses only an occasional property which is not common to the whole class, the nominal group to which it applies cannot have variable denotation. Therefore, (16) is fine, but (17) is ill-formed.

(16) *(Jakákoli) multičára se skládá ze dvou a více paralelních čar.*
(Any) MULTILINE consists of two and more parallel lines.

(17) * *Jakákoli multičára spojuje tyto dva body. (Any multiline connects the two points.)*

So we saw that variable determination can occur in clauses which express as a whole variable denotation. Besides that, nominal groups with variable determination can also occur

in clauses whose modality is other than «reality», e.g. probability, usuality, ability, obligation, and also in interrogative and imperative clauses and clauses in future tense. This means that in the AGILE instructions we can always express variable determination if needed if we want the text to be explicit.

At this point we have finished the discussion of Hlavsa's modes of determination and the means to express them in Czech. We compared them with Halliday's partial non-specific deictics. To complete our comparison with Halliday's classification of deictics, we have to turn our attention also to some applicators that Hlavsa discusses under the heading of quantification, namely those corresponding to the English each, every, all, no and both. In other words, total non-specific deictics.

Partial variable deictic items (Czech)	
selective (unmarked for knowledge of individual characteristic)	*nějaký, jakýkoli (any)*
non-selective (marked for lack of knowledge of individual characteristic)	*některý, kterýkoli (any)*

Figure 47 Partial non-specific variable deictic items in Czech

2.8.4.4.1 Evidence from the Czech Corpus

There are no occurrences of nominal groups with explicit existential deictics in the Czech corpus, even though a number of nominal groups with no explicit deictic item could be classified as realizing existential determination.

2.8.4.4.2 Formalization of Partial Variable Deictics in Czech

In the following table, we propose a formalization of the partial variable determination in Czech on the basis of the above discussion. The variable deictics *nějaký* and *některý* have been already reflected in Figure 41

```
VARIABLE-PARTIAL-SELECTIVE-TYPE
(variable, partial, elective) → [JAKYKOLI] (Deictic ! JAKYKOLI)

VARIABLE-NONSELECTIVE-TYPE
(variable, partial, nonselective) → [KTERYKOLI] (Deictic ! KTERYKOLI)
```

Figure 48 Formalization of existential non-specific determination

2.8.4.5 Partial Variable determination in Russian

Variable determination in Russian is again quite similar to Czech. The primary realisation of the deictic element in this function is *какой-нибудь*. Since Russian does not have the variety of jaký vs. který with respect to the knowledge of an individual characteristic (selectivity), there are no further distinctions among the possible the realisations. Besides *какой-то (any)*, variable determination can also be expressed by *любой (arbitrary)*.

2.8.4.5.1 Evidence from the Russian Corpus

There are no occurrences of nominal groups with explicit existential deictics in the Russian corpus.

2.8.4.5.2 Formalization of Partial Variable Deictics in Russian

In the following table, we propose a formalization of the partial variable determination in Russian on the basis of the above discussion.

```
VARIABLE-PARTIAL-SELECTIVE-TYPE
(variable, partial, selective) → [LJUBOJ] (Deictic ! LJUBOJ)

VARIABLE-NONSELECTIVE-TYPE
(variable, partial, nonselective) → [VSJAKIJ] (Deictic ! VSJAKIJ)
```

Figure 49 Formalization of existential non-specific determination

2.8.5 Total Non-Specific Determination

As noted earlier, Halliday draws a distinction between totality and partiality when classifying non-specifying deictic elements.

All the determiners we discussed in the preceding sections on unique, existential and variable determination in fact signal quantity: (i) they are distinguished for number: the singular forms can be used with singular or mass nouns (quantity one), and the plural ones with plural nouns (indefinite quantity); (ii) they all are characterized as partial. Note also that in Un or Ex determination, singular number really means denotation of a single entity, while in the case of Var, there can be more then one entities denoted.

We are now concerned with total non-specific determination. This mode of determination means that the entire set of things referred to by the nominal group or all the things in this set are concerned.

2.8.5.1 Total Non-Specific Detemination in Czech

The primary applicators realizing totality in Czech are *všechen* (*all*, mass nouns), *všichni* (*all*, count nouns), *oba* (*both*), *každý* (*every*), *žádný* (*none*), *celý* (*entire*). Their summary is shown in Figure 50.

Total non-specific deictics (Czech)			
Polarity	**Collectivity**	**count**	**mass**
positive	individual	*každý(every)* *celý (entire)*	*všechen (all)*
	unmarked	*všichni (all)* *oba (both)*	
negative	unmarked	*žádný (none)*	

Figure 50 Total non-specific deictic items in Czech

2.8.5.1.1 Evidence from the Czech Corpus

The Czech corpus contains several occurrences of non-specific total Deictic items. They are all positive and combined with count nouns. We list them in the table in Figure 51.

Nominal groups with a total quantifying applicator	
každý segment křivky (34)	*each line segment*
relativně ke všem ostatním nakresleným elementům (87)	*relative to all elements drawn so far*
každý element (105)	*every element*
vlastnosti celé multičáry (106)	*properties of the entire multiline*
úsek nebo oblouk pro každou multičáru (110)	*segment or arc for each multiline*
konec celé multičáry (112)	*the end of the entire multiline*
výpis všech typů (157)	*listing of all types*
všechny čáry. (284)	*all lines*

Figure 51 Occurrences of nominal groups with an explicit total non-specific deictic

2.8.5.1.2 Formalization of Total Non-Specific Deictics in Czech

In the following table, we propose a formalization of the total non-specific determination in Czech on the basis of the above discussion.

```
TOTAL-TYPE
(total) → [positive-total]   [negative-total]

POSITIVE-TYPE
(positive-total countable) → [individual]  [non-individual]

VSECHEN
(positive-total noncountable) → Deictic ! VSECHEN

ZADNY
(negative-total) → Deictic ! ZADNY

INDIVIDUAL-TYPE
   [KAZDY] (Deictic ! KAZDY)
   [CELY] (Deictic ! CELY)

NONINDIVIDUAL-TYPE
(non-individual) →
   [more-then-two] (Deictic ! VSICHNI)
   [dual] (Deictic ! OBA)
```

Figure 52 Formalization of total non-specific determination

2.8.5.2 Total Non-Specific Detemination in Russian

The situation is like in Czech.

2.8.5.2.1 Evidence from the Russian Corpus

The following table overviews the occurences of total non-specific deictics in the Russian AGILE corpus.

Nominal groups with a total quantifying applicator	
для каждого конца мультилинии (88)	*for each end of multiline*
чтобы показать список всех стилей (120)	*to list all styles*
все несохраненные линии (196)	*all unrecorded lines*

Figure 53 Occurrences of nominal groups with an explicit total non-specific deictic

2.8.5.2.2 Formalization of Total Non-Specific Deictics in Russian

In the following table, we propose a formalization of the total non-specific determination in Russian on the basis of the above discussion.

```
TOTAL-TYPE
(total) → [positive-total]   [negative-total]

POSITIVE-TYPE
(positive-total countable) → [individual]  [non-individual]

VESJ
(positive-total noncountable) → Deictic ! VESJ

NIKAKOJ
(negative-total) → Deictic ! NIKAKOJ

INDIVIDUAL-TYPE
   [KAZHDYJ] (Deictic ! KAZHDYJ)
   [TSELYJ] (Deictic ! TSELYJ)

NONINDIVIDUAL-TYPE
(non-individual) →
   [more-then-two] (Deictic ! VSE)
   [dual] (Deictic ! OBA)
```

Figure 54 Formalization of total non-specific determination

2.8.5.3 Total Non-Specific Detemination in Bulgarian

Total Non-Specific Deictics (Bulgarian)		
Polarity	**Number**	
	Singular	**Plural**
positive	*Всеки (each)*	*всички (all)*
negative	*Никой (neither)*	*никои (none)*

Figure 55. Total Non-Specific Deictics in Bulgarian

In Bulgarian there are no different items for total non-specific deictics in relation to «Collectivity».
Deictic items in singular can be marked for gender.

2.8.5.3.1 Evidence from the Bulgarian Corpus

The only occurrences of total non-specific deictic items in Bulgarian corpus are three:

всеки сегмент на полилинията (28)	*each polyline segment*
всеки край на мултилинията (84)	*each end of the multiline*
За да видите списък на всички стилове (122)	*To list all styles*

2.8.5.3.2 Formalization of Total Non-Specific Deictics in Bulgarian

```
TOTAL-POLARITY
(total) → [positive-total]   [negative-total]

TOTAL-NUMBER
(total) → [singular]   [plural]

VSEKI
(positive-total singular) → Deictic ! VSEKI

VSICHKI
(positive-total plural) → Deictic ! VSICHKI

NIKOI-S
(negative-total singular) → Deictic ! NIKOI-S

NIKOI-P
(negative-total plural) → Deictic ! NIKOI-P
```

Figure 56. Formalization of total non-specific determination in Bulgarian

2.8.6 Concluding Remarks

We have discussed DEICTIC elements in Czech, Bulgarian and Russian. Both Czech and Russian differ from English in that they do not have articles, neither definite nor indefinite. Bulgarian has a definite article, which is attached as a suffix. In Czech and Russian, nominal groups do not have to contain a DEICTIC element unless the speaker wants to make specificity or non-specificity explicit. In Bulgarian, specificity needs to be marked by the definite article, while non-specific nominal groups do not have to be accompanied by a Deictic element.

In all three languages, like in English, it is possible to use demonstratives or adjectives as **determiners**, i.e. to realise the DEICTIC elements to express specificity or non-specificity of the set denoted by the nominal group. Besides the DEICTIC element, we also encounter the equivalents of the English POST-DEICTIC; these are realised by adjectives. Since the realisations of the Deictic element are grammatically adjectives or pronouns, they have to

agree with their head noun in number and gender, and in Czech and Russian also in case. The Bulgarian definite suffix is also varied for number and gender.

We have noted that possessive pronouns and other possessives easily occur in nominal groups which have an explicit Deictic element in all three languages. Moreover, the possessive pronoun itself does not make a nominal group specific. Therefore, we proposed to include treat possessives along with other attributes within a nominal group, rather than as a type of Deictic elements.

The discussion of the meaning and realisation of Deictic elements in Czech and Russian has distinguished between three modes of determination, namely unique, existential and variable, where unique determination corresponds to specific determination in the SFG terminology and existential and variable determination are subtypes of non-specific determination in SFG. This more fine-grained classification proved useful. There are determiners in our languages that are used specifically to realise the Deictic element in one of these types of determination. There are also a few non-specific determiners which are less specialized and can be used for both existential and variable determination.

In Bulgarian, there does not appear to be determiners used specifically for either Existential or Variable determination. Therefore, the treatment of non-specific determination in Bulgarian follows closely the treatment of determination in the grammar of English and does not make this finer-grained distinction.

We also discussed the close interaction between determination and quantification. In all the three languages, there are distinct forms of determiners with regard to number. In this way, the determiners in fact quantify as well. In singular number they even specify the quantity precisely, in plural number they determine a quantity larger than one. As we discussed above, non-specific determination can either span over the entire set of denoted entities (total) or a subset thereof (partial). The interplay between determination and quantification is important for choosing appropriate realisations.

The overall summary of our analysis of determination in the three languages is that in the cases when explicit determiners are used, the situation is quite similar to English and it is not very difficult to adapt the systemic networks for our purposes. In particular, the choices are based on the same features as in English.

The question when to insert an explicit Deictic element and when not is interesting and not trrivial at all in the Slavonic languages. As noted above, nominal groups in Czech and Russian and also non-specific nominal groups in Bulgarian do not have to contain an explicit Deictic element. Such is possible when specificity or non-specificity can be deduced by the hearer from the context or from general knowledge, or it is also possible that the speaker does not make it explicit because it is not important for the content he is communicating. We also hinted above at certain heuristics that correlate information structure and determination. In particular, in Czech and Russian, there is a tendency to interpret contextually bound nominal groups as definite. However, this is not more than a tendency.

At any rate, whether or not to include an explicit Deictic element is a decision that should be made in the course of text planning; so, in the context of the AGILE generation system, by the text structurer. The reason for this is that the decision needs to be based, among other things, on a model of the context in which the entities referred to so far are represented. We shall therefore return to these issues in the context of the WP5 work package.

3. Conclusions and Further Work

The research reported here has addressed the following phenomena: Transitivity, Diathesis, Mood, Tense, Aspect, Clause complexity, Word order and Determination in nominal groups. With respect to Transitivity and Diathesis, the Slavonic languages under consideration are similar in their way of structuring the piece of reality reported in a clause in terms of process, participants, circumstances – which classes also closely follow the Upper Model. In Czech and Russian the system of morphological case is used quite heavily for distinguishing types of processes and participants. In such general semantic functions as location or causation the languages are similar too, in the sense that they are realised by "universal" syntactic means as prepositional phrase or conjunctions, though these languages often have some additional means to convey the same oppositions as English: location, direction, orientation. Also, since all Slavonic languages have more developed inflectional system, they show greater rigidity in transition between parts of speech in comparison to English, so morphological alternations are necessary for nominalization (cf. English *drawing a line* vs. Russian *рисование линий*) and classification resources (*a dialogue box* vs. *диалоговое окно*; *a toolbar* vs. *панель инструментов*); this complicates the task of lexical choice.

Since Russian and Czech have the obligatory case specification of a nominal group and belong to the nominative language type, they have a marked realisation of the Actor / Goal opposition, in which the Goal is treated as a participant undergoing a manipulation by the Actor. So prototypically in the active voice the direct complement is realised by the accusative case and the subject - by the nominative case, however, sometimes (in particular, in Czech in our project), some verbs which have the ideational Goal role still are not considered as transitive ones, since the Goal is not expressed by the accusative case (for example, *odpovídat na otázky*, *answer to questions*). Also, since all Slavonic languages have more developed inflectional system, they show greater rigidity in transition between parts of speech in comparison to English, so morphological alternations are necessary for nominalization (cf. English *drawing a line* vs. Russian *рисование линий*) and classification resources (*a dialogue box* vs. *диалоговое окно*; *toolbar* vs. *панель инструментов*); this complicates tasks of lexical choice.

Even though Bulgarian belongs to the family of Slavonic languages, its recent developments have happened in the Balkan Languages Union, which includes such genealogically diverse languages as South Slavonic (Bulgarian, Serbo-Croat), Romance (Romanian), Greek and Albanian languages, which share significant amount of syntactic features due to convergence phenomena (for example, suffixation of an element expressing definiteness in a nominal group). So on a scale from analytic to synthetic, Bulgarian is much more analytic than Czech or Russian: it (as English) only has rudimentary case markers and distinguishes participant roles by their positions and prepositions accompanying them, rather than by morphological cases. Correlating with the lack of case is the fact that word order is not as flexible as in Czech or Russian.

While several choices in the Mood system of Slavonic languages are possible for expression of instructions (for example, clauses in imperative mood, infinitive clauses, impersonal indicatives, and impersonal modal assertions), our corpus favours use of imperatives continuing the English tradition, since it has been translated from the English CAD manual. With respect to MODALITY our corpora offer no instances of occurrence of modal adjuncts. However, modal auxiliaries are used in Slavonic languages, in particular, in the modal sense of possibility (*можно)* and volition (*хотеть*). The verb 'want/wish'

(*хотеть*) is only used once in a subordinate clause (*нажимайте* ***r*** *каждый раз, когда хотите …)*. This is significantly different from the situation in the Bulgarian corpus where the equivalent verb *искам* occurs much more frequently.

The Slavonic languages under consideration share the distinction between perfective and imperfective aspects as a lexical characteristic of the verb. Thus, there are no significant differences in our corpora, either. Both types of verb aspect are used with a preference for the perfective aspect in the clauses in the imperative mood (e.g., in Russian *введите, укажите)* and the perfective aspect in declarative, descriptive clauses (e.g., in Russian, *позволяет рисовать, задает, создает*). Occurrences of the imperfective aspect in the imperative mood are rare, in fact there appars to be only one at the same point of the text as in the Bulgarian corpus, in Russian, «*нажимайте* ***r*** *наждый раз, когда хотите …»*.

Since the register of software manuals belongs to registers of science and technology, which favour use of complex syntactic structures, a very wide range of clause-complexity relations is present in the corpus (for example, practically all types of expansion are found in our corpus). The clause-complexity typology initially developed in SFG for English happened to be sufficient to code the relations we find in our languages, while their realisations often differ in their structure. In particular, the Slavonic languages, and especially Russian, often use non-restrictive participle modifiers for hypotactic elaboration. With respect to expression of hypotactic enhancement, spatial and temporal relations are realised in similar ways. Simultaneity in Russian is expressed by adverbial-participial clauses, which are obsolete in Czech and absent in Bulgarian. The most variation across languages is found in relations expressing Manner and Purpose. Since texts of this register in all our languages are mostly translated, they often follow the English pattern, which sometimes result in marginally grammatical or stylistically inappropriate sentences; in particular, purpose is often expressed by prepositional groups using nominalisation of the process (*Pro ukončení úsečky stiskněte ENTER, Для завершения рисования линии нажмите Enter, To end the line, press Enter*).

The fact that Slavonic languages exhibit free word order is widely acknowledged. However, in our research we tried to identify reasons underlying alternations in word order using notions of Topic-Focus articulation developed in the Prague School paradigm (Functional Generative Description, FGD). The above-mentioned fact that participants of Bulgarian clauses are marked by their position instead of case, leads to restrictions for alternating word-order constructions which are grammatical in Bulgarian. However, in contrast to English it allows relative freedom for order of the elements following the verb. The default order in Czech and Russian follows the information structure, in which the verb marks the border between context-bound elements which are tended to be put in the first position and non-bound elements which are put last.

In the course of this research we also developed a comparison of approaches to the text structure in SFG and FGD. Expression of the deictic element in nominal groups has been the topic of our research in determination (word-order constraints on determination are in the section on the word order). In all three languages, like in English, it is possible to use demonstratives or adjectives as determiners, i.e. to realise the DEICTIC elements to express specificity or non-specificity of the set denoted by the nominal group. An important result of comparison is that the three Slavonic languages have distinguished between *three* modes of determination, namely unique, existential and variable, where unique determination corresponds to specific determination in the SFG terminology and existential and variable determination are subtypes of non-specific determination in SFG.

The question of when to insert an explicit Deictic element and when not is interesting and not trivial at all in the Slavonic languages. As noted above, nominal groups in Czech and Russian and also non-specific nominal groups in Bulgarian do not have to contain an explicit Deictic element. This is possible when specificity or non-specificity can be deduced by the hearer from the context or from general knowledge, or it is also possible that the speaker does not make it explicit because it is not important for the content he is communicating. We also hinted above at certain heuristics that correlate information structure and determination. In particular, in Czech and Russian, there is a tendency to interpret contextually bound nominal groups as definite.

The accomplishments of Task WP6.2 reported in this deliverable cover quite substantial part of grammatical description for Bulgarian, Czech and Russian. The functional tradition is widely developed in Slavonic linguistics, in particular, Jakobson and Prague School studies should be mentioned, however, there exist very little resources which help in formal specification of functional phenomena in these languages. Hence the work presented here contributes substantially towards the systemic-functional description of Slavonic languages. This work goes beyond the straight needs of modelling our existing corpus, however. As stated in the introduction, the project goals require a balanced specification of grammars of the target languages. This specification can be specialised later for a particular domain and text style to be generated.

WP6.2 serves the basis for a computational implementation of respective grammars to be developed in the course of WP7.2 providing grammatical resources for the intermediate demonstrator. The formal notation used for specification of phenomena allows easy rendering into KPML statements for system definitions. Additional specifications to grammatical classification resources described here consist in definition of a formal interface (choosers and inquiries for systems) and testing completeness and coverage of produced systemic choice network. Since the grammatical classifications developed in WP6.2 have wider coverage than necessary for the intermediate demonstrator, only a part of them will be implemented in WP7.2. These classifications will also serve to identify any adjustments necessary to the Upper Model to accommodate these three languages.

In addition to the above-mentioned goals of WP6.2 which are internal for AGILE, the resources described in this document are useful for other researchers whose main focus is on related aspects of natural language processing (e.g., natural language understanding, machine translation and automatic abstracting systems) and crosslingual information retrieval. Thus this project provides resources for other project for languages of Eastern Europe. As far as we are aware, outside of machine translation, no generation systems have yet been developed for these languages.

References

[CORP] Anthony Hartley, Danail Dochev, Nevena Gromova, Kamenka Staykova, Alla Bemova, Aleksandr Rosen, Jiri Trojanek, Elena Sokolova. *Tagging and analysis of instructional texts in the software domain.* June 1998. The deliverable for the WP3 of AGILE project PL961104.

[LSPEC2] Geert-Jan M. Kruijff, John Bateman, Alla Bémová, Danail Dochev, Ivana Kruijff-Korbayová, Serge Sharoff, Hana Skoumalová, Lena Sokolova, Kamena Stoikova, Elke Teich, Jiří Trojánek. *Modelling Lexical Resources in KPML for Generating Instructions in Slavonic Languages.* October 1998. The deliverable for the WP4-2 of AGILE project PL961104.

[MODL1] Richard Power, *Preliminary model of the CAD/CAM domain.* June 1998. The deliverable for the WP2-1 of AGILE project PL961104.

[SPEC1] John Bateman, Elke Teich, Ivana Kruijff-Korbayova, Hana Skoumalova, Nevenna Gromova, Danail Dochev, Serge Sharoff, Elena Sokolova, *Specification of grammatical resources for the initial demonstrator.* June 1998. The deliverable for the WP6-1 of AGILE project PL961104.

AutoCAD (1995) *AutoCAD Release 13. User's guide,* chapter 2. Autodesk Co.

Bateman J.A. (1992) Grammar, systemic. In Stuart Shapiro (ed.), *Encyclopedia of Artificial Intelligence,* Second Edition, pages 583–592. John Wiley and Sons, Inc.

Bateman, John, Kasper, Robert, Moore, Joanna, and Whitney, Richard, (1990) *A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model*. Technical Report, University of Southern California/Information Sciences Institute: Marina Del Rey CA.

Daneš František [1970] Zur linguistischen Analyse der Textstruktur. Folia linguistica, No. 4, pp. 72-78.

Daneš František [1985] Věta a text [A sentence and a text]. Studie a práce lingvistické, Vol. 21.Prague.

Daneš F. et al. 1987. Mluvnice češtiny, Vol. 3. [Czech Grammar]. Academia.

Firbas Jan [1992] Functional sentence perspective in written and spoken communication. Cambridge: Cambridge University Press.

Hajičová Eva [1993] Issues of sentence structure and discourse patterns. Theoretical and computational linguistics, Vol. 2. Prague: Charles University.

Hajičová Eva [1995] Surface and underlying word order. In: Prague Linguistic Circle Papers, Vol.1, Amsterdam: John Benjamins, 113-124.

Hajičová Eva and Korbayová Ivana [1998] On the notion of topic. Prague Linguistic Circle Papers. John Benjamins Publishers. Amsterdam, the Netherlands. in press.

Hajičová Eva and Sgall Petr [1987] The Ordering Principle. Journal of Pragmatics. 11(4): pp. 435-454.

Hajičová Eva and Vrbová Jarka [1981] On the salience of the elements of the stock of shared knowledge. Folia linguistica 65: 291-303.

Hajičová Eva and Vrbová Jarka [1982] On the role of the hierarchy of activation in the process of natural language understanding. In: Proceedings of Coling 1982 (ed. by J. Horecký], Prague: Academia, pp. 107-113.

Halliday M.A.K. [1967]. Notes on transitivity and theme in English. Journal of Linguistics 3, 1967, 37-81, 199-244; 4, 1968, 179-215.

Halliday, M.A.K. [1970] Language structure and language function. In: New Horizons in Linguistics. John Lyons (ed.) Penguin Books Ltd. pp. 140-165.

Halliday, M.A.K. (1973) *Explorations in the Functions of Language.* Edward Arnold, London.

Halliday, M.A.K. (1978) *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M.A.K. (1985) *Introduction to Functional Grammar,* London: Edward Arnold.

Halliday M.A.K. (1994). An Introduction to Functional Grammar. Second Edition. Arnold.

Halliday M.A.K. and Hasan R. (1976). Cohesion in English. Longman.

Hartley, A., Paris C. (1995) *French corpus analysis and grammatical description*. DRAFTER Deliverable WP14/15, ITRI, University of Brighton, UK.

Henschel, R., Bateman, J.A., (1997) Application-driven automatic subgrammar extraction. In: Proceedings of ACL/EACL97 Workshop: "ENVGRAM: Computational Environments for grammar development and linguistic engineering".

Hlavsa Z. 1975. Denotace objektu a její prostředky v současné češtině [Denotation of an Object and its Means in Contemporary Czech]. Studie a práce lingvistické, Vol. 10. Academia

Karlík Petr, Marek Nekula, Zdenka Rusínová (eds.) 1995. Příruční mluvnice češtiny [Handbook of Czech Grammar]. Nakladatelství Lidové Noviny.

Krifka Manfred [1992] A compositional semantics for multiple focus constructions. Linguistische Berichte, Sonderheft Informationsstruktur und Grammatik, ed. by J. Jacobs, pp. 17-53.

Křížková H. 1971. Systém neurčitých zájmen v současných slovanských jazycích [The System of Non-Specific Pronouns in Contemporary Slavonic Languages]. Slavia 40, p. 342-370.

Kruijff Geert-Jan [1998] Basic Dependency-Based Logical Grammar. Technical Report, UFAL-1998-05.

Kruijff-Korbayová Ivana [1998] The Dynamic Potenital of Topic and Focus: An Approach to Discourse Representation Theory. PhD Thesis. Charles University, Prague.

Kruijffová Ivana and Hajičová Eva [1997] Topics and centers: A comparison of the TFA approach and the centering theory. Prague Bulletin of Mathematical Linguistics. 67:25-50.

Mathesius Vilém [1936] On some problems of the systematic analysis of grammar. In: Travaux du Circle Linguistique Prague, Vol. 6. John Benjamins Publishers.

Mathesius Vilém [1939] O takzvaném aktuálním členění větném [On the so-called functional sentence perspective]. Slovo a slovesnost 5: 171-174.

Matthiessen, C., Bateman J. (1992) *Text Generation and Systemic Functional Linguistics: Experiences from English and Japanese.* London: Pinter Publishers.

Matthiessen, C. (1996) *Lexicogrammatical Cartography of English.* Singapore.

Nasr, A., Rambow, O., Kittredge, R., (1998) A linguistic framework for controlled language systems In: Proceedings 2nd International Workshop on Controlled Language Applications (CLAW'98), Pittsburgh PA, 19-22 May 1998.

Partee Barbara H., Hajičová Eva and Sgall Petr [1998] Topic-Focus Articulation, Tripartite Structures and Semantic Content. Kluwer Academic Publishers.

Petkevič Vladimír [1995] A new formal specification of underlying structures. In: Theoretical linguistics, 21(1): 7—61. Walter de Gruyter.

Poldauf Ivan. 1969. The so-called mediopassive in English. AUC-Philologica 3, Prague Studies in English 13:15-34.

Poldauf I. and Šprunk B. 1968. Čeština – jazyk cizí [Czech – a foreign language] Principles of Structural syntax I. Reader MFF Charles University, Prague.

Sgall Petr [1967] Functional sentence perspective in a generative description. Prague studies in mathematical linguistics, Vol. 2: 203-225. Prague.

Sgall Petr and Hajičová Eva [1987] The ordering principle. Journal of pragmatics 11: 435-454.

Sgall Petr, Hajičová Eva and Benešová Eva [1973] Topic, focus and generative semantics. Kronberg. Taunus.

Sgall Petr, Hajičová Eva and Buráňová Eva [1980] Aktuální členění věty v češtině (Topic-focus articulation of the Czech sentence). Academia. Prague, Czech Republic.

Sgall Petr, Hajičová Eva and Panevová Jarmila [1986] The meaning of the sentence in its pragmatic aspects. Mey, J.L. (ed.), Reidel.

Sgall P., Pfeiffer O. E., Dressler W. U and Půček M. 1995. Experimental Research on Systemic Ordering. Theoretical Linguistics, 21(2/3): 197-239.

Šmilauer Vladimír. 1947. Novočeská skladba [Modern Czech Grammar]. Published by ing. Mikuta, Prague.

Šmilauer, V. (1972) *Nauka o českém jazyku.* SPN, Praha.

Steedman Mark [1991] Structure and Intonation. Language, 68. pp. 260-296.

Steiner E., Teich E., (in press) German: a metafunctional profile. In: Caffarel A. & Martin J.R, *Functional Typology*, Benjamins, Amsterdam.

Vallduví Enric [1992] The informational Component. Garland, New York.

Weil Henri [1844] translation in Weil [1887].

Weil Henri [1887] The Order of Words in the Ancient Languages compared with that of the Modern Languages. Translated by Super, Charles W. Boston: Ginn & Company, Publishers. New edition with an introduction by Aldo Scaglione. John Benjamins b.v. (1978). Amsterdam, The Netherlands.

Андрейчин Л., Попов К., Стоянов С. (1977) *Граматика на българския език.* София: Наука и изкуство.

Бояджиев, Т., Куцаров, И., Пенчев, Й. (1998) *Съвременен български език*. София: "Петър Берон".

Золотова, Г.А. (1988) *Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса.* (*Syntactic dictionary: an inventory of elementary features of the Russian syntax*). Москва: Наука.

Падучева, Е.В. (1996) *Семантические исследования: Семантика времени и вида в русском языке; Семантика нарратива*. (Semantic studies: semantics of time and aspect in Russian; semantics of narratives). Москва: «Языки русской культуры».

Стоянов, С. (1980) *Граматика на българския книжовен език*. София: Наука и изкуство.

Храковский, В.С. (1996) Грамматические категории глагола (опыт теории взаимодействия). In: *Межкатегориальные связи в грамматике*. СПб: Дмитрий Буланин. (in Russian)

Циммерлинг, А. (1998) Синтаксическая деривация и диатезные преобразования. In: *Proceedings of the International workshop on computational linguistics and its applications Dialog'98* Vol.1. Казань: Хетер. (in Russian)

Шведова, Н.Ю., (ред.) (1970) *Грамматика современного русского языка.* (*Modern Russian grammar*). Москва: Наука. (in Russian)

4.Appendix

It includes a sample text in English, Bulgarian, Czech and Russian adapted from the AutoCAD manual (p. 45 in the original English edition for AutoCAD v.13)

English
To create a multiline style
First open the Multiline Styles dialog box using one of these methods:
Windows From the Object Properties toolbar or the Data menu, choose Multiline Style.
DOS and UNIX From the Data menu, choose Multiline style.

1. Choose Element Properties to add elements to the style.
2. In the Element Properties dialog box, enter the offset of the first line element.
3. Select Add to add the element.
4. Choose Color. Then select the element's color from the Select Color dialog box.
5. Choose Linetype. Then select the element's linetype from the Select Linetype dialog box.
6. Repeat these steps to define another element.
7. Choose OK to save the style of the multiline element and to exit the Element Properties dialog box.

Bulgarian
Създаване стил на мултилиния
Отначало отворете диалоговия прозорец с име Multiline Styles, като използвате един от следните методи:
Windows От функционалния ред с име Object Properties или менюто Data изберете Multiline Style.
DOS и UNIX От менюто Data изберете Multiline Style.

1. Изберете Element Properties, за да прибавите елементи към стила.
2. В диалоговия прозорец Element Properties въведете отместването на елемента на мултилинията.
3. Изберете Add, за да добавите елемента.
4. Изберете Color. След това изберете цвета на елемента от диалоговия прозорец Select Color.
5. Изберете Linetype. След това изберете вида линия на елемента от диалоговия прозорец Select Linetype.
6. Повторете стъпки 2-5, за да дефинирате друг елемент.
7. Изберете OK, за да запишете характеристиките на елемента на мултилинията и излезете от диалоговия прозорец Element Properties.

Czech
Vytvoření stylu multičáry
Pro vytvoření stylu multičáry
nejdříve otevřete dialogový panel Styly multičár jednou z následujících metod:

Windows Z nástrojového panelu Vlastnosti objektů nebo z menu Data vyberte *Styl mutičáry*.
DOS a UNIX Z menu Data vyberte *Styl multičáry.*
1. Vyberte *Vlastnosti prvků* pro přidání elementů ke stylu.
2. V dialogovém panelu Vlastnosti prvků zadejte posunutí elementu multičáry.
3. Vyberte *Přidat* pro přidání elementu.
4. Vyberte *Barva*. Poté zvolte barvu elementu z dialogového panelu Výběr barvy. Vyberte *Barva*. Objeví se dialogový panel Výběr barvy. Zvolte barvu elementu.
5. Zvolte *Typ čáry*. Poté zvolte typ čáry daného elementu z dialogového panelu Výběr typů čar.
   Zvolte *Typ čáry*. Objeví se dialogový panel Výběr typů čar. Zvolte typ čáry daného elementu.
6. Pro vytvoření dalšího elementu opakujte tyto kroky.
7. Vyberte OK pro uložení vlastností elementu multičáry a pro opuštění dialogového panelu Vlastnosti multičáry.

Russian
Чтобы создать стиль мультилинии
Сначала откройте диалоговое окно Multiline Styles одним из следующих способов:
Windows В панели инструментов Object Properties или в меню Data выберите пункт Multiline Style.
DOS & UNIX В меню Data выберите пункт Multiline Style.

1. Нажмите кнопку Element Properties, чтобы добавить элементы в стиль.
2. В диалоговом окне Element Properties введите смещение первого элемента линии.
3. Нажмите кнопку Add, чтобы добавить этот элемент.
4. Выберите пункт Color. Затем выберите цвет элемента в диалоговом окне Select Color.
5. Выберите пункт Linetype. Затем выберите тип линии элемента в диалоговом окне Select Linetype.
6. Повторите эти шаги, чтобы задать еще один элемент.
7. Нажмите кнопку OK, чтобы сохранить стиль этого элемента мультилинии и закрыть диалоговое окно Element Properties.