AGILE

Automatic Generation of Instructions in Languages of Eastern Europe

---

Title ***Specification of grammatical resources for the Initial Demonstrator***

Authors John Bateman
Elke Teich
Ivana Kruijff-Korbayova
Hana Skoumalova
Nevenna Gromova
Danail Dochev
Serge Sharoff
Elena Sokolova

Deliverables *SPEC1, SPEC1-Bu, SPEC1-Cz, SPEC1-Ru*

Status *Final*

Availability *Public*

Date *June 1998*

INCO COPERNICUS PL961104

Abstract:

This document contains the deliverables SPEC1-Bu, SPEC1-Cz and SPEC1-Ru of Work Package 6, Task 6.1 of the AGILE project. We present the scope of grammar modelling in the target texts, describe the theoretical background of the approach taken (Systemic Functional Linguistics, SFL) and introduce the notation that is used for grammar specification. This is followed by discursive descriptions of those phenomena present in the target texts as well as their formal specification in the SFL notation. The specification is concluded with cross-linguistic analysis and considerations for further activity.

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
fax:	+44-1273-642908

Table of Content

1. Introduction 1
   1.1 Overview of principles of Systemic Functional Linguistics 1
   1.2 Notational conventions 3
   1.3 Approach to grammar development 4
2. Language-specific reports 6
   2.1 Bulgarian (SPEC1-Bu) 6
      2.1.1 Clause structure. 6
      2.1.2 Structure of nominal groups 6
      2.1.3 Structure of verbal groups. 7
      2.1.4 Structure of prepositional phrases 7
   2.2 Czech (SPEC1-Cz) 8
      2.2.1 Nominal groups 8
      2.2.2 Sentence structure for complex SPL's 8
      2.2.3 Speech act: Imperative 9
      2.2.4 Transition between categories 9
      2.2.5 Agenthood in a pro-drop language 9
   2.3 Russian (SPEC1-Ru) 10
      2.3.1 Clause structure 10
      2.3.2 Structure of nominal-like groups 12
      2.3.3 Structure of verbal groups 14
      2.3.4 Structure of prepositional phrases 14
3. Conclusion 14
4. Appendix 17
   4.1 Target text 1 (adapted from the AutoCAD manual, p. 45) 17
      4.1.1 English 17
      4.1.2 Bulgarian 17
      4.1.3 Czech 17
      4.1.4 Russian 17
   4.2 Target text 2b (adapted from the AutoCAD manual, p. 49-50) 17
      4.2.1 English 17
      4.2.2 Bulgarian 18
      4.2.3 Czech 18
      4.2.4 Russian 19

1. Introduction

The research described in this document has been devoted to specification of linguistic phenomena found in target texts for the Initial Demonstrator in Bulgarian, Czech and Russian (source texts in these three languages and their English equivalent are shown in the appendix). These texts present simple procedures for achieving a single goal, as in the following sentences in these languages:

Стартирайте командата MLINE като изберете Multiline от плаващото меню на функционалния ред с име Draw.

Spustíme příkaz MLINE vybráním Multičára z nabídky Multičára v nástrojích Kreslení.

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.

In this research we adopted the following assumptions:

- the functional approach in terms of systemic-functional linguistics (an overview of SFL is given Section 1.1);
- a formal specification of phenomena found in these texts using the systemic notation (the notation is presented in Section 1.2);
- a contrastive description based on resources provided by the existing English grammar (the principles of this investigation are exposed in Section 1.3)

1.1 Overview of principles of Systemic Functional Linguistics

The linguistic-theoretical basis of the project and the KPML system used for its implementation is Systemic Functional Linguistics (SFL). SFL is a British school of linguistic thought (Halliday 1973), belonging to the tradition of functional approaches to language (Hjelmslev 1943), (Jakobson 1971), (Dik 1978), (Bondarko 1984) and showing affinities with the Continental-European Prague School (Firbas 1966), (Daneš 1974), (Sgall et al 1986). The two basic theoretical postulates of SFL according to (Halliday 1978) are:

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

These postulates are also shared by the above-mentioned theories, however, in contrast to them SFL provides a formal descriptive tool called Systemic Functional Grammar (SFG)[1]. SFG is functional in that it acknowledges three main functions around which languages are organized: the ideational, the interpersonal and the textual (see below). SFG is systemic in that the main focus in description is on the grammatical paradigm (or: system). The grammar of a language is represented as a *system network*, which can be read as a declarative statement of grammatical features and the co-occurrence constraints holding between them. An SFG is thus a classification-based approach to grammar, rather than a rule-based one---very similar to other models of grammar currently used in computational linguistics, such as

[1] For a fairly concise introduction to Systemic Functional Grammar see (Bateman 1992). More extensive accounts of the theory can be found in (Halliday 1985), (Halliday and Matthiessen, to appear).

Head-Driven Phrase Structure grammar (Pollard & Sag 1994), in which a classification hierarchy of grammatical (and lexical) types constitutes the grammatical description. What makes SFG stand out from such approaches is the functional motivation of grammatical types. The grammatical types are functionally motivated in the following way.

The notion of function in SFL is predominantly manifested in the concept of *meta-functions*, a set of generalized functions that language is said to fulfil. The ideational meta-function encodes a language's propositional content. Its grammatical aspect 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, Spatial Adjunct, Temporal Adjunct etc. The interpersonal meta-function encodes speakers' roles in an interaction, their attitudes and evaluations. One of the major grammatical reflexes is the clause systems of mood, which distinguishes between declarative, interrogative and imperative and accounts for the differences in syntactic structure that these different moods come along with. The textual metafunction encodes properties of textual organization, such as global text structure, coherence and cohesion. In the grammar, this is reflected in systems of theme-rheme patterning and information structuring at clause level and determination at the level of the nominal group. The functionally motivated systems and their features are associated with *realization statements*. The realization statements are the attributes of a functional grammatical class and specify the syntagmatic, surface-syntactic constraints that the functional classes exhibit. For instance, a surface-syntactic constraint associated with the functional class declarative of finite clauses in English is that in syntactic structure the Subject is ordered before the Finite verb. This distinction between paradigmatic, functionally motivated classes and syntagmatic structure is referred to as *axiality*: a linguistic description in SFG always has these two aspects that hang together by the relation of realization.

Another organizing principle of linguistic representation in SFG is *rank*. The top system in the grammatical classification is the rank system. The rank system distinguishes between clauses, nominal groups, prepositional phrases, adjectival and adverbial groups, words and sometimes morphemes. This *rank scale* gives the basic paradigmatic grammatical classes for which particular sets of systems and their features hold and it defines the basic constituency organization of syntagmatic structure.

At the rank of a simple clause all three metafunctions are expressed in different structures:

1. The ideational metafunctions is described as a configuration of Process (for example, Material-Creative: *to draw a line*), Participants (which functions also depend on the type of Process) and Circumstances. Classification of processes, participants and circumstances are provided by the Upper Model (Bateman et al 1996).
2. The structures of the interpersonal metafunction are related to binding of the content of a clause to interaction between persons involved in the speech act; this includes choices of mood, modality, tense, etc.
3. Since the basic meaning of the textual metafunction is contextualisation of a message in text, it provides cohesion of a narrative text by controlling the choice of theme and rheme structures, topic-focus articulation, ellipsis, choice of expressions referring to objects and actions, i.e. such phenomena as (in)definiteness, pronominalisation, synonyms, etc.

1.2 Notational conventions

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

Notation element	Example	Comments
functional elements	Actor, Subject, etc	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
realization statements:		Statements which constrain ways in which features are manifested in language
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, in the Russian system of MOOD types such instructions as:
Укажите конечную точку полилинии. (*Specify the endpoint of the polyline.*)

are described by

```
MOOD [imperative : polite] (Nonfinitive:::imperative, plural)
```

meaning that in such sentences a choice of imperative is followed by a default choice of its polite form, and the realisation statement constrains inflectional characteristics of Nonfinitive verbal constituents by the features of imperative and plural.

1.3 Approach to grammar development

The whole grammar of particular language in SFG consists of a paradigmatic description of linguistic phenomena (with reasons for their choices) and a syntagmatic description of their realisation. This inherently supports development of multilingual grammar, since a paradigmatic classification in SFG terms is based on classification of communicative needs which are often generalised across languages. For example, at the clause rank in the range of meanings of the interpersonal metafunction there is a paradigmatic choice between giving and demanding information, the latter choice leads to choosing between confirmation of the proposition involved and identification of an object (or a meta-object). This classification is logically grounded in the structure of communication in the society, so it is natural that it is presented in any language. This paradigmatic description (semantics of grammatical features) is shared across languages, while syntagmatically they are realised in different ways (by the word order, morphology, etc), for example, in English the first choice is realised by declarative and interrogative sentences which differ in presence and position of an auxiliary verb. Since the existing English grammar is the most advanced systemic description developed so far (it has about 1200 systems), it is used as a resource for paradigmatic classification of Bulgarian, Czech and Russian phenomena found in the Initial Demonstrator texts.

Since the grammar specifications of the target languages for the Initial Demonstrator are based on the existing English grammar, we proceed in inductive fashion: sentences in Bulgarian, Czech and Russian are coded using English grammar in order to detect grammatical phenomena which differ from features of the English grammar both in their syntagmatic realisation and paradigmatic classification. The majority of alternations concern alternation of realisation statements (for example, realisation of imperative), some classifications are altered or created (for example, the system of cases in Russian, which should be supported by semantic reasons for choosing case features). This document describes alternations in the English grammar which are necessary to take into account phenomena of Slavonic languages for texts of the Initial Demonstrator only.

The complete semantic specification for a given sentence or phrase to be generated provides environment for the grammar. Said environment in our case consists of statements in Sentence Plan Language (SPL, (Kasper 1989)). The most important part of SPL expressions is taken from the Upper Model concepts and relations (Bateman et al. 1996); this is the resort of ideational (i.e., propositional) meaning in SFG. Besides the ideational information expressed in an spl, also interpersonal information (e.g., :speechact command) and textual information (e.g., :identifiability-q identifiable) is contained. The non-ideational kinds of information typically come from text planning.

Draw a polyline.

```
(S / DISPOSITIVE-MATERIAL-ACTION :LEX DRAW
   :SPEECHACT COMMAND
   :ACTEE (D / OBJECT :LEX POLYLINE))
```

Drawing of a polyline.

```
(S / DISPOSITIVE-MATERIAL-ACTION :LEX DRAW
   :EXIST-SPEECH-ACT-Q NOSPEECHACT
   :ACTEE (D / OBJECT :LEX POLYLINE))
```

Figure 1. Examples of SPL expressions

Here it is necessary to point out a difference between SPL expressions (as a language-dependent encoding of semantics) and grammatical functions as they are used by language. Often there is a one-to-one mapping between the both (for example, :speechact command is routinely realised by the imperative mood), sometimes there is a many-to-many mapping (as in the case of :identifiability-q identifiable). In many cases such a mapping depends on external conditions. For example, two SPL expressions shown in Figure 1 are identical with the exception of that the first one is realised by an imperative sentence and the second one by a nominal group, so :actee of the first SPL is realised as Goal conflated with function of Direct-complement, while the second :actee is a qualifier of a nominal group (grammatical terms follow (Halliday 1985)).

In the case of a *semantic* difference in expression of the same content in English and Slavonic languages, two solution methods are possible for modelling correct target language sentences on the basis of the English systemic network:

- technical solutions - local idiosyncratic alternations of the SPL expressions in order to achieve a correct target sentence;
- systemic solutions - an extension of Upper Model to add concepts and roles relevant to Slavonic languages, though keeping closely to the English model, if at all possible.

In the case of a technical solution SPL is used not for input of semantic data, but rather for encoding surface phenomena. In a systemic solution SPL keywords encode meanings of functions, which are used in target languages; by devising these language-dependent meanings we also try to reach some generalizations across target languages. The difference between two types of solutions can be illustrated by the following example. Complex notions in English are typically expressed by a class ascription, for example, an SPL expression for a *control system*:

```
(o / object :lex system :class-ascription (c / abstraction :lex control))
```

while in Russian this notion is expressed as *система управления* using the genitive case. The genitive case in itself in Russian has a range of meanings often coinciding with the meaning of the English preposition *of*, which is encoded in SPL by means of the `part-of` semantic relation from the Upper Model. So a technical solution in our specification would use the following SPL:

```
(o / object :lex sistema :part-of (c / abstraction :lex upravlenie))
```

and the following realisation of `part-of` (in English the Pselector is realised by *of*):

```
[partitive] (+Part) (Part:nominal-group, genitive)
   (+Pselector) (Pselector!ellipsiszero)
```

However, this misses the fact that *control* is by no means a part of *system*, so such a realisation is absolutely incorrect from the semantic ground, though this choice allows a rapid prototyping for a limited number of texts (as it is in the case of the Initial

Demonstrator). A systemic solution for this construction would consist in investigation of different functions of the genitive case in Russian in order to reflect this in an updated Upper Model. For example, the range of meanings encoded by the genitive case in Russian includes relations of part-whole (*точка линии - a point of a line*), class-class restriction (*система управления - control system*), object-owner (*кнопка мыши - mouse button*), attribute-object (*цвет заливки - background colour*), container-containment (*чашка молока - a cup of milk*), event-time of this event (*демонстрация первого мая - demonstration at the first of May*). So this can be captured in a representation that some concept fills a slot of another concept, and subtypes of this filling relation are part-of, superclass, owner, attribute, location and time. An analysis of other ways for expressing these relations and possible corrections in the Upper Model are given in Section 0 below.

2. Language-specific reports

2.1 Bulgarian (SPEC1-Bu)

2.1.1 Clause structure.

There are no important differences in the functional structure. Since most clauses of the target texts for the Initial Demonstrator are imperative they do not need an explicit Subject.

The complex clauses in the Initial Demonstrator texts are mainly hypotactic:

purpose

The purpose is expressed in Bulgarian by means of a subordinate finite clause (the verb agrees with the Subject in gender and number), the subordinate clause is introduced by a hypotactic conjunction "*za da*" ("in order to").

manner

The expression of manner in the Bulgarian text slightly differs from that in the English one, because the verb in the subordinate clause is again finite (agrees in person and number with the Subject). The hypotactic conjunction is "*kato*".

2.1.2 Structure of nominal groups

The experiential structure of nominal groups in Bulgarian is very similar to the English one.

Deictic	Numerative	Epithets	Classifier	Thing

We have observed one main difference concerning the order of these functional elements, i. e. the **Classifiers** in Bulgarian can appear both before and after the Thing. Nominal Classifiers follow the Thing, which implies some changes in the KPML systems.

Classifiers which are expressed by Adjectives in Bulgarian have to precede the noun, like the Epithets.

Deictic	Numerative	Epithets	Classifier Adjective	Thing	Classifier Noun

Another major difference consists in the lack of agreement in gender and number between Adjectives and ordering **Numeratives** and the Noun in English and the obligatory agreement between those elements in Bulgarian.

The identifiability of the noun is expressed in English by a specific, determinative Deictic **the** which is a separate English word found in the lexicon**,** whereas in Bulgarian it is

expressed by means of morphemes attached to the noun. If the noun is modified, then the morpheme is attached to the first constituent of the nominal group. In our case the Adjectives expressing Quality and the ordering Numeratives will have to take the definite article morpheme, that morpheme has separate forms for masculine, feminine, neuter gender and one single form for plural for the three genders. For masculine it has a full and short form. The choice between the two forms depends on the function of the constituent in the sentence. In the cases when English grammar uses a non-specific, non-selective, partial, singular Deictic **a(n),** in Bulgarian no article is needed.

Examples:

*parva**ta** tochka na poliliniata* (**the** first point of the polyline)

*krajna**ta** tochka na multiliniata* (**the** last point of the multiline)

chertane na multilinia (drawing of **a** multiline)

Some of the Bulgarian nominal groups had to be further elaborated and the preposition **of** (*na*) has to be inserted in order to achieve a more natural looking Bulgarian phrase. In this case the Qualifiers are expressed by prepositional phrases:

Deictic	Numerative	Epithets	Classifier	Thing	Qualifier

Examples:

stila na multiliniata (the style of the multiline)

mastaba na multiliniata (the scale of the multiline)

spisaka na vidovete linii (the list of line types)

imeto na stila (the name of the style)

tekstovijat prozorec na AutoCAD (AutoCAD Text Window)

2.1.3 Structure of verbal groups.

All the verbs occurring in the Initial Demonstrator texts express directed material processes.

The interpersonal function is expressed by mood. The mood of the sentences occurring in the Initial Demonstrator texts is mainly imperative. Only one verb form is reflexive and Indicative in 3 p. sg (*se pojavjava*). Bulgarian imperative has two forms: one for 2 p. sg. and one for 2 p. pl. In Bulgarian instructional texts the imperative used is in polite form, i.e. 2 p. pl. of the verb. The Imperative in English is non-finite verb form so the English default is Subject implicit. On the other hand, the English requires for a finite verb an explicit Subject. Bulgarian does not need necessarily a Subject for a finite verb form, because the verb has an ending, which shows the person and number of the omitted Subject. In imperative clauses it is more natural to have an implicit Subject.

2.1.4 Structure of prepositional phrases

The types of prepositional phrases in the sentences of the Initial Demonstrator are very limited. Most of them function as Adjuncts of the circumstantial type in the clause. All circumstantial Adjuncts are of the spatial- location type.

ot plavashtoto menu Polyline (from the Polyline flyout)

na funktsionalnija red Draw, na ekrana (from the Draw toolbar, on the screen)

The preposition depends strongly on the preceding verb.

vavejdam v komandnija red (enter at the prompt)

2.2 Czech (SPEC1-Cz)

The solutions chosen for the Initial Demonstrator as described here tend to be pragmatically based and do not attempt to make any unrecoverable decisions; we are aware that the Initial Demonstrator was meant to serve as a test whether the project partners have mastered the tools available and how universal the tools are; the material we have gathered in this stage will be exploited in the intermediate stage of the project.

As an example of a pragmatical solution, the realization of SPL units for concepts as lexical items can be given. Most of the T-box concepts are specified in such a way that they are directly mapped onto lexical units. This might cause difficulties in the model for Czech as in the other languages; however, all the instances found in the test corpus can be solved by inserting the Czech equivalent into the dictionary.

2.2.1 Nominal groups

The issue encountered most frequently is that of the inner structure of the nominal groups in case of 'object: class-ascription': e.g. PLINE command *příkaz PLINE* (lit. 'command PLINE'), Polyline layout *nabídka Polyline* (lit. 'layout Polyline'), Draw toolbar *nástroje Kreslení* (lit. 'tools Draw'). This, again, can be considered as a matter of grammar: the preferential choice of surface structure for class-ascription (if the category of the modifier is a Name) is to modify the head noun (in the above examples command, layout, toolbar, respectively) by a postposited modifier.

2.2.2 Sentence structure for complex SPL's

Our analysis of the corpus indicates that in case of simple SPL's (e.g. those of the type *directed action*) no difficulties are involved and the structure of the Czech sentence (if the appropriate topic-focus articulation is observed, see below) can be more or less directly mapped from the SPL structure. However, in case of complex SPL's (e.g. those of the types rst-purpose, rst-means manner), the situation is not so simple.

Let us illustrate this point by some examples from the corpus.

(i) The (English) target sentence is:

Start the PLINE command by choosing Polyline from the Polyline flyout on the Draw toolbar.

This sentence is an introductory sentence for a whole set of instructions how to draw a polyline. In Czech, it is more natural to split the complex sentence into two parts and to take the first part as a kind of 'headline' for the whole set. The English literal paraphrase of the Czech counterpart thus is:

'We start the PLINE command: from the Polyline flyout on the Draw toolbar we choose Polyline.'

The same situation obtains in the set of commands for drawing a multiline. In both cases, the SPL is of the rst-means manner type, with the action: start as the domain and the action: choose as the range.

(ii) The same sentence may serve as an example of the necessity to introduce into the text structuring procedure an account of topic-focus articulation of the sentence, at least in a simplified form. In the English target sentence, the 'source' (from the Polyline flyout on the Draw toolbar) is given by the context, it is contextually bound and part of the topic (in the Praguian sense), while the 'actee' of 'choose' (Polyline) is the focus. In the unmarked, prototypical case, topic precedes focus in the Czech sentence; therefore, the Czech surface order is as exemplified in the literal paraphrase above ('from the Polyline flyout on the Draw toolbar we choose Polyline').

Another example of a similar sort is the English target sentence

The AutoCAD text Window appears.

Here, the focus (in the Praguian sense) is The AutoCAD text Window; the action of appearance is contextually bound. The Czech counterpart (in its literal E. paraphrase) is, as expected, with the focus at the end:

'Appears the AutoCAD text window.'

2.2.3 Speech act: Imperative

In all instances of speech act: imperative, the Czech corpus uses also the imperative mood. In Czech the 1st person plural of the indicative mood is usally used in manuals, which, however, does not hold for the analysed corpus. In fact, there is a considerable variation in expressing instructions across various types of manuals, depending to a large extent on the manufacturer's or distributor's preferences.

2.2.4 Transition between categories

One of the distinguishing feature of Czech technical style, in contrast to the general tendency of Czech to use verbal rather than nominal constructions, is a frequent use of nominalizations. Noun groups are frequently used where English often uses non-finite clauses (with gerunds and infinitives). In the present corpus, nouns are used in the full grammatical sense - having all morphological (inflection) and syntactic properties of nouns. Thus e.g. the English target sentence

Press Enter to end the multiline

Zmáčkněte Enter pro ukončeni multičáry.

In its literal translation into English:

'Press Enter for the end (Noun) of the multiline.'

In such cases, the SPL directed-action has a (verbal) noun rather than a verb as its lex; it will be a matter of the grammatical choice to select a construction with a (deverbal) noun.

2.2.5 Agenthood in a pro-drop language

Czech verbs have a rich morphology. There are different finite forms also for the first, second and third person of each number. In the past participle, the third person both singular and plural also reflects differences in gender. Moreover, Czech as well as Bulgarian and Russian distinguishes between the "formal" (polite) and "informal" second person (which is apparent in singular number). All these possibilities are applicable in the declarative mood. However, not all of them are applicable in instructions. In general, instructions can be formulated using the informal or formal imperative of second person singular or plural, or

using the first person plural. The latter form can be seen as referring to both the writer of the manual and its reader (user of the system). But this form is also conceived of as realizing a "general actor". Therefore, the first person plural can be considered, together with the imperative form, the most neutral form of reference, suitable for use in the instructions in the AGILE texts.

With respect to the system of agency in SFG, the sentences in the first person plural realizing the general actor appear to share some characteristics of both agency-specified and agency-generic. On the one hand, the voice is effective and active, which corresponds to agency-specified, on the other hand, no particular actor needs to be assumed. Czech has another possibility for expressing agency-generic by using a form usually called reflexive passive, e.g. *Zadá se koncový bod* ('One enters the end point' or 'The end point is entered'). The verb is in third person singular, and the reflexive particle *se* marks the structure as not requiring the subject.

We have not attempted to solve these issues in a principled manner in the current phase of the project, but they should be addressed in the next phase. In our current teatment the use of first person plural is specified in the SPL, regardless of whether it realizes a general (not specified) actor or a specified actor including the writer and the reader.

Czech is a so-called pro-drop language. In general, subject is dropped when the entity referred to is easily recoverable from the context. The finite verb form usually carries sufficient information to distinguish between the different grammatical persons. If, however, a personal or demonstrative pronoun is used as subject, it is interpreted contrastively. It occurs either as non-thematic (part of the rheme), or as thematic, especially in cases where subject of the preceding sentence refers to another entity.

In non-imperative instructions, a dropped subject corresponds predominantly to the first person plural, which is then considered to refer to the writer and the reader (or the general actor as mentioned above). Subject dropping also occurs with the third person singular or plural in the Autocad texts. These are then cases of reference to objects of the user interface, commands or objects created by the user.

In order to achieve a satisfactory handling of subject dropping in Czech in the course of text generation in AGILE, we need to develop such a strategic generation which specifies, on the one hand, whether a full nominal form of a referring expression should be used, or whether an abbreviated (pronominal) form is sufficient, and, on the other hand, provides enough information to enable the decision whether the pronoun can be dropped. Then we need to define in the tactical generator the systems and choosers that drop the Subject. We will address these issues in the next phase of the project.

2.3 Russian (SPEC1-Ru)

2.3.1 Clause structure

The basic ideational structure of clauses is the same as in English: Process + Participants + Circumstances. Classes of the semantic input for these functions in SPL are taken from the Upper Model. In the Initial Demonstrator all Processes are of the class dispositive-material-process, Participants involved in the process have the role of Actee; Circumstances are either :source or :spatial-locating.

The range of interpersonal functions in clauses of the Initial Demonstrator is restricted to imperative and indicative, expressed by corresponding verb forms. Predicates in all clauses in

our target texts have a direct complement which is realised by a nominal group in the accusative case. By default a direct complement is located after a verb. The most frequent choice of the imperative mood leads to the grammatical realisation in simple sentences without an explicit subject:

Укажите конечную точку полилинии. (*Specify the endpoint of the polyline.*)

The functional structure of this clause is:

```
[SENTENCE]
     [VOICE#1/NONFINITIVE#1/LEXVERB#1/PROCESS#1]..="Ukazhite"
     [DIRECTOCOMPLEMENT#1/GOAL#1/MEDIUM#1]
     [DEICTIC#2]
     [STATUS#2]
         [QUALITY#3] ..="konechnuju"
     [THING#2] ..="tochku"
     [PORTION#2]
         [MINORPROCESS#4]
         [MINIRANGE#4]
             [DEICTIC#5]
             [THING#5] ..="polilinii"
```

The source SPL form is:

```
(E / DIRECTED-ACTION :LEX UKAZATJ :SPEECHACT IMPERATIVE
    :ACTEE (D / OBJECT :LEX TOCHKA
       :PROPERTY-ASCRIPTION  (O1 / QUALITY :LEX KONECHNYJ)
       :GENERAL-GENITIVE  (P / OBJECT :LEX POLILINIYA)))
```

Two possible ways for expression of imperative, polite and non-polite, are realized in Russian by using plural and singular forms of the verb respectively. The default choice in texts for the Initial Demonstrator is polite.

This means that choices at the clause rank in the Russian grammar involve:

```
MOOD [imperative : polite] (Nonfinitive:::imperative, plural)
EFFECTIVE-VOICE [operative] (Directcomplement:accusative)
```

In indicative sentences (there is only one example in the target texts for the Initial Demonstrator) the function of Subject is realised by the nominative case.

In principle other types of clauses may have other functions of participants realised by other cases, in particular, the functions of Addressee, Receiver and Client are typically realised by the dative case; Manner or Effective Agent in non-Subject role - by the instrumental case. As for other cases, the genitive case is typically used inside the nominal group structure for recursive nesting of qualifiers; in the clause rank it marks Subject in an existential negation (***Его*** *не было дома* - ***He*** *was not at home*).

```
[negative-indicative & existential] (Subject:::genitive)
```

The prepositional case is used only in prepositional phrases. In texts of the Initial Demonstrator it expresses only locative Circumstance functions.

There are three cases which complicate this situation:

1. logical or temporal coordination between two actions, as in:

Введите “j” и выберите выравнивание

Enter j and choose a justification

This case is already taken into account in the English grammar under paratactic extension type sentences. The only alternation in the Russian grammar consists in lexicalisation of a conjunction, for example:

```
QUALIFYING-COORDINATION-TYPE [Temporal-Coordination]
      (COORDINATOR ! ZATEM)
```

2. expression of a purpose of an action, as in:

В строке команд введите "st", чтобы выбрать стиль.

Enter st at the prompt to select a style.

The purpose is expressed by an infinitival clause, which function is a hypotactic elaboration of the main clause. Subordinator *чтобы* is used for conjunction of these two clauses. In formal notation:

```
CIRCUMSTANTIAL-DEPENDENT[Purpose-dependent](Subordinator!chtoby)
ENHANCING-FINITENESS[Nonfinite-enhancing](Nonfinitive:::infinitive)
```

The locative circumstance in this sentence is positioned in the thematic position of the sentence in order to leave the last position for the focal actee. This should be taken into account by the text structuring component.

3. expression of means used for achieving a goal.

Запустите команду MLINE, выбрав Multiline...

Start the MLINE command by choosing Multiline ...

This is realised by an adverbial participial group, which function is a hypotactic elaboration of the main clause. This group is separated from the main clause by a comma. The functional structure of an adverbial participial group is also realized at the clause rank in the same way as in rank-shifted clauses:

```
MANNER-CONDITION-TYPE [Means-dependent] (Nonfinitive::adv-participle)
```

2.3.2 Structure of nominal-like groups

The basic functional structure of nominal groups in Russian is the same as in English (according to Halliday (1985), Section 6.2):

Deictic	Numerative	$Epithet_1$	$Epithet_2$	Classifier	Thing	Qualifier
Те	*многие*	*красивые*	*старые*	*электрические*	*поезда*	*с пантографами*
Those	*many*	*splendid*	*old*	*electric*	*trains*	*with pantographs*

All nominal groups function in a clause as participants which are objects, some of which are classified according to their dimensions (for example, one-or-two-d-location).

The most notable difference from English is agreement between Thing, which is expressed by a noun, and functions from Deictic to Classifier which are expressed by adjectival words and agree with Thing in case, number, gender and animacity. In target texts for the Initial Demonstrator there is no need to model the animacity agreement (as in *рисовать красного коня—draw a red horse* vs. *рисовать красный квадрат—draw a red square*), since all objects in our texts are inanimate. Gender is specified in the lexical item for Thing, number is assigned by information which is provided by the environment.

One of six cases (Nominative, Genitive, Dative, Accusative, Intstrumental, Prepositional) is preselected for a nominal group at the clause or prepositional group rank. Actual assignment of this case to Thing, the head of the nominal group, is achieved in the pronoun-

case system at the nominal-group rank (the name of the group should be changed in the future). Then the case feature is copied to all its elements, when they are inserted.

The deictic element in the nominal group structure is expressed by possessive or demonstrative pronouns; there is no explicit tool in Russian such as an article for expression of definiteness, usually the deictic status of the nominal group is not expressed at all or expressed by word order. By this reason, *a/the* realisation of Deictic is replaced, for example, in:

```
INDIVIDUAL-DETERMINATION (Deictic!ellipsiszero)
```

There are complex cases in processing nominal groups with numeratives in the nominative case: a quantitative numerative has the same case as preselected in clause systems (this is also true for deictics, which are placed before the numeratives), but the rest of the nominal group is in the genitive case; the number of adjectives is plural, and number of the noun, which serves as Thing of the nominal group, is either plural if the numerative is relatively large (greater than or equal to 5) or singular otherwise:

эти	*три*	*новых*	*дома*
these (pl, nom)	*three (nom)*	*new (pl, gen)*	*house (sing, gen)*
эти	*пять*	*новых*	*домов*
these (pl, nom)	*five (nom)*	*new (pl, gen)*	*houses (pl, gen)*

There are five functional structures of Russian nominal groups relating Thing to Classifier. Instead of a single role (:class-ascription used for this goal in English), the Function of Classifier may be realised in Russian using semantic roles:

1. Label-ascription: *MLINE command — команда MLINE*.
2. Index-ascription: *page 1 — страница 1* (a numerical index can be expressed in another way by ordinatives: *первая страница / first page*)
3. General-genitive: *style list — список стилей* (note an additional plural number in Russian, this should be taken into account in lexicon that list operates with multiple values par excellence, so the text structuring application which generates SPL using information from the domain A-box should insert :multiplicity-q multiple).
4. Class-ascription proper: *stone wall — стена из камня* (but the fifth structure is more often: *каменная стена*).
5. Property-ascription: *electric kettle — электрический чайник*

For grammar modeling, the most interesting distinction in corresponding grammatical constructions is between attribution of qualities and substances. This is realised by an adjective in a pre-position to the Thing and by a noun in the genitive case in a post-position, respectively. From the practical viewpoint it is easier in the Russian grammar to split substances and qualities (merged in the fourth case) into different semantic roles: Substance-class-ascription realised by a qualifying noun in the genitive case (*усилитель мощности - power amplifier*); Quality-class-ascription realised by an adjective in the function of Classifier (*каменная стена*); Material-class-ascription realised by a qualifying noun in the genitive case with a corresponding preposition (*стена из камня*). The General-genitive role is useful for expression of different types of genitive functions. In the English grammar the range of meanings of the genitive case is coded by Part-of; in the Russian grammar the role of General-genitive has been made synonymous to Part-of to keep the Upper Model

intact (see above the discussion about the difference between technical and systemic solutions):

```
[partitive] (+Part) (Part:nominal-group, genitive)
   (+Pselector) (Pselector!ellipsiszero)
```

The ascription of a label in the Russian grammar is modeled by ordering the label after the end of the nominal group (in order to take into account the case, when a nominal group has qualifiers in the genitive case *панель инструментов Draw — Draw toolbar*):

```
CLASSIFICATION-TYPE [labeled] (Thing...Classifier1)
```

2.3.3 Structure of verbal groups

In the texts for the Initial Demonstrator there is no internal structure of verbal groups: they consist simply of verbs in the perfective aspect and imperative morphological form: the first feature is ascribed by the lexicon, the second one is preselected by the choice of imperative in the system of mood. The structure of a verbal group can be complicated by additional operators, such as modality, polarity, reflexivity (the latter is realised morphologically in the verb); there are no secondary tenses in Russian verbal groups, though these elements are not presented in the Initial Demonstrator texts.

2.3.4 Structure of prepositional phrases

All prepositional phrases in target texts for the Initial Demonstrator realise Locative Circumstances. So their minor ranges are preselected with the prepositional case. Realisation of their minor-processes depends on the Circumstance role, semantic and syntactic features of the Process and the class of a corresponding object. There are three cases in our examples:

:spatial-locating + object + *выбрать*: ***в*** *палитре Polyline* (*choose* ***from*** *the Polyline flyout*)

:spatial-locating + one-or-two-d-location: ***на*** *панели инструментов Draw* (***on*** *the Draw toolbar*)

:spatial-locating + zero-d-location + *введите*: ***в*** *строке команд* (*enter* ***at*** *the prompt*)

3. Conclusion

In the course of this work package we achieved the grammatical specifications for the sample texts in Bulgarian, Czech and Russian (see sample texts in the appendix). This has only been possible because we could make use of the strategy of sharing paradigmatic resources. The English grammar in KPML, NIGEL, as well as the specification of English given in (Halliday 1985) and (Matthiessen 1995) have been used as the basis for resource sharing. English is a Germanic language and therefore typologically different from Slavonic languages, such as Bulgarian, Czech and Russian, but since the main body of linguistic description in Systemic Functional Grammar is functional and resource sharing operates primarily on functional categories, this approach has been successful. Also, as an observation from linguistic specification and implementation, even though Bulgarian belongs to the family of Slavonic languages, on a scale from analytic to synthetic, Bulgarian is much closer to English than Czech or Russian: It only has rudimentary case and syntactic agreement and it operates with prepositions rather than with morphological case. Correlating with the lack of case is the fact that word order is not as flexible as in Czech or Russian. Also since Slavonic languages have more developed inflectional system, they show greater rigidity in transition between parts of speech in comparison to English.

So far, we cannot claim that we have general specifications of the grammars of Bulgarian, Czech and Russian. However, the methodological steps from here are clear and the results we have achieved so far will easily scale up. While we have so far just changed existing systems very locally, for instance, PRIMARY-CLASSIFICATION in all three languages, MOOD for imperative in Czech and Russian and for declarative in Czech, in the next phase of implementation, we are going to deal with whole paradigms and work according to functional regions (transitivity, circumstantials, qualification, classification etc). The grammatical complexity of the texts occurring in the kinds of manuals we deal with from the CAD-CAM domain will not be much higher than already covered, so that we can concentrate on developing more general solutions. Also, while resource sharing has worked quite well generally, there are some issues to be dealt with that relate to the typological particularities of Bulgarian, Czech and Russian as Slavonic languages. Among these are the following. On the technical side, for Czech and Bulgarian external morphology components will be linked up with KPML, just as it has been done for Russian. On the linguistic modelling side, we have to develop a treatment for word order, which is highly variable in Russian and Czech and almost entirely cotextually and contextually determined (topic-focus articulation, information distribution). Also, in all of the three languages there are some very complex agreement phenomena, in particular, in the nominal group, that need a principled treatment.

The experiences in this first phase have shown that it is more than worthwhile to work contrastively across Bulgarian, Czech and Russian and possibly distribute the descriptive and specificational efforts, thus not only sharing resources with English, but also among the three languages under investigation in this project. The results of this phase of work in work package 6 also feed into other work packages, most notably grammatical resource implementation (work package 7), which involves expression of paradigmatic choices and their realisation in terms of KPML system (Bateman et al. 1996).

Our further work in this Work Package 6 and the closely related Work Package 7 assumes making existing descriptions more “universal” with respect to any specific language and target Slavonic languages in general (as an example, cf. specification of quality and substance class ascriptions in Russian in Section 0). The system of syntactic functions is also related to expression of rhetorical relations, for example, the rhetorical relation of precondition in German and Russian is expressed by means of a prepositional phrase with preositions *bei* and *ïðè* respectively (*Bei fehlerhaften Anschluß des Gerätes, При неправильном включении устройства; Bei stark kalkhaltigem Wasser, При наличии в воде...*). English lacks such a construction, a precondition is expressed by either subordinate clause using *when, unless* (*when tap water is hard*), or in another system of functions (*Any error in connecting the appliance*). Hence, this implies an interrelation of this activity with Work Package 5 (Text structuring)

References

Bateman, J., Henschel R., Rinaldi F. (1996) *The Generalized Upper Model 2.0.* URL: *http://www.darmstadt.gmd.de/publish/komet/gen-um/newUM.html*

Bateman J.A. (1992) Grammar, systemic. In Stuart Shapiro (ed.), *Encyclopedia of Artificial Intelligence,* Second Edition, pages 583–592. John Wiley and Sons, Inc.

Bondarko A.V. (1987) *Theory of Functional Grammar,* Leningrad: Nauka.

Daneš F., (1974) *Papers on Functional Sentence Perspective.* Academia, Prague.

Dik S., (1978) *Functional Grammar.* North Holland, Amsterdam.

Firbas J., (1966) On defining the theme in functional sentence analysis. *Travaux Linquistique de Prague,* 1:267--280.

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., Matthiessen C.M.I.M., (to appear) *Construing Experience through Meaning: a Language-based Approach to Cognition*.

Hjelmslev L., (1943) *Omkring sprogteoriens grundlaeggelse.* Akademisk Forlag, Køpenhavn.

Jakobson R., (1971). *Selected Writings*. The Hague: Mouton.

Kasper, R.T. (1989) A flexible interface for linking applications to PENMAN's sentence generator. In *Proceedings of the DARPA Workshop on Speech and Natural Language.* Available from USC/Information Sciences Institute, Marina del Rey, CA.

Matthiessen, C.M.I.M (1995) *Lexicogrammatical cartography.* Tokyo, Taipei and Dallas, International Language Sciences Publishers.

Matthiessen, C.M.I.M, Bateman J., (1992) *Text Generation and Systemic Functional Linguistics: Experiences from English and Japanese.* London: Pinter Publishers.

Pollard C.J., Sag I.A., (1994) *Head-driven Phrase Structure Grammar.* Chicago University Press, Chicago.

Sgall P., Hajicová E., and Panevová J. (1986) *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* Reidel Publishing Company, Dordrecht.

Sharoff S., Sokolova L., (1998) Towards ML-generation of manuals: the initial stage of the AGILE project. *Proc. International workshop on computational linguistics DIALOG'98*, October, 1998.

4.

Appendix

4.1 Target text 1 (adapted from the AutoCAD manual, p. 45)

4.1.1 English

To draw a polyline

1. Start the PLINE command by choosing Polyline from the Polyline flyout on the Draw toolbar.
2. Specify the first point of the polyline
3. Specify the endpoint of the polyline.
4. Press Return to end the polyline.

4.1.2 Bulgarian

Чертане на полилиния

1. Стартирайте командата PLINE, като изберете Polyline от плаващото меню на функционалния ред с име Draw.
2. Задайте началната точка на полилинията.
3. Задайте крайната точка на полилинията.
4. Натиснете Return, за да завършите полилинията.

4.1.3 Czech

Nakreslení křivky.

1. Spustíme příkaz KŘIVKA vybráním Křivka z nabídky Křivka v nástrojích Kreslení.
2. Určíme počáteční bodkřivky.
3. Určíme koncový bod křivky.
4. Křivku dokončíme stisknutím Enter.

4.1.4 Russian

Чтобы нарисовать полилинию / Рисование полилиний

1. Запустите команду PLINE, выбрав пункт Polyline в палитре Polyline на панели инструментов Draw.
2. Укажите первую точку полилинии.
3. Укажите конечную точку полилинии.
4. Нажмите клавишу Return, чтобы завершить рисование полилинии.

4.2 Target text 2b (adapted from the AutoCAD manual, p. 49-50)

4.2.1 English

To draw a multiline

1. Start the MLINE command by choosing Multiline from the Polyline flyout on the Draw toolbar.
2. Enter **st** at the prompt, then enter **?** to select a style.
   The AutoCAD Text Window appears.
   Enter the name of the style.
3. Enter **j**, then enter a justification from top, zero and bottom to change the justification of the multiline.
4. Enter **s**, then enter a number to change the scale of the multiline.
5. Close the AutoCAD Text Window.
6. Specify the first point of the multiline.
7. Specify the second point of the multiline.
8. Specify the third point of the multiline.
9. Press Return to end the multiline.

4.2.2 Bulgarian

Чертане на мултилиния.

1. Стартирайте командата MLINE като изберете Multiline от плаващото меню на функционалния ред с име Draw.
2. Въведете st след съобщението за избор на вида линия.
3. Въведете ?, за да се появи на екрана списъка с видовете линии.
4. Въведете j за изравняване на линията и изберете горно, нулево или долно изравняване.
5. Въведете s и мащаб, за да промените мащаба на мултилинията.
6. Сега задайте мултилинията.
7. Задайте началната точка на мултилинията.
8. Задайте втората точка на мултилинията.
9. Задайте третата точка на мултилинията.
10. Натиснете Return, за да завършите мултилинията.

4.2.3 Czech

Nakreslení multiáry

1. Spustíme příkaz MČÁRA vybráním Multičára z nabídky Multičára v nástrojích Kreslení.

2. Zadáme **st** na příkazovém řádku.

3. Vybereme druh čáry.

   Napíšeme **?.**

   Objeví se dialogové okno AutoCad Text Window.

   Zadáme druh čáry.

4. Změníme zarovnání čáry.

   Napíšeme **j**.

   Zadáme zarovnání vzhledem k hornímu nebo dolnímu okraji nebo ke středu.

5. Změníme měřítko čáry

   Napíšeme **s**

   Zadáme měřítko.

6. Zavřeme dialogové okno AutoCad Text Window.
7. Určíme první bod čáry.
8. Určíme druhý bod čáry.
9. Určíme třetí bod čáry.
10. Čáru dokončíme stisknutím Enter.

4.2.4 Russian

Чтобы нарисовать мультилинию / Рисование мультилиний

1. Запустите команду MLINE, выбрав Multiline в палитре Polyline на панели инструментов Draw.
2. В строке команд введите st, затем введите ?, чтобы выбрать стиль.

На экране появится окно AutoCAD Text Window.

Введите имя стиля.

3. Введите j, затем выберите выравнивание сверху, по центру и снизу, чтобы выровнять мультилинию.
4. Введите s, затем введите масштаб, чтобы изменить масштаб мультилинии.
5. Закройте окно AutoCAD Text Window.
6. Укажите первую точку мультилинии.
7. Укажите вторую точку мультилинии.
8. Укажите третью точку мультилинии.
9. Нажмите клавишу Return, чтобы завершить рисование мультилинии.