

Syntactic functions of infinitive in contemporary Czech¹

Vladimír Petkevič
Institute of Theoretical and Computational Linguistics
Faculty of Arts
Charles University
Vladimir.Petkevic@ff.cuni.cz

1. Introduction

This paper is a case study that deals with syntactic functions of infinitive in contemporary Czech using large textual data – extensive corpora of contemporary Czech. If every important syntactic and morphological phenomenon in contemporary written Czech is thoroughly described on the basis of huge repositories of data represented by corpora of contemporary written Czech, then automatic tools for morphological disambiguation and syntactic analysis can considerably be improved. It is exactly the lack of linguistic knowledge that lies behind errors in morphological disambiguation (both statistical and rule-based one) as well as in parsing of contemporary written Czech (both statistical and rule-based one). The study wants to contribute to the improvement of morphological disambiguation and syntactic analysis of infinitives by formulating a set of statements identifying as precisely as possible conditions for the identification of syntactic functions of infinitive in a sentence. These statements can be implemented as formal rules used in morphological taggers and parsers of Czech (e.g. within the project *Syntactic Annotation of Corpora*, cf. grant GAČR No. P406/10/0434); moreover, the results of statistical parsers of Czech (especially McDonald's MST parser, cf. Novák & Žabokrtský 2007) can be corrected by these rules implemented in automatic correction software tools.

In the study, the following corpora of contemporary Czech were used:

- SYN2010 (representative, 100 mil. word forms; part-of-speech (POS) and morphologically tagged)
- a working treebank SYNT (48 mil. word forms, POS and morphologically tagged; marked also with syntactic functions and syntactic governors by McDonald's stochastic parser).

The paper is organized as follows: first, the identification of ambiguous infinitival forms as a preliminary step for syntactic analysis will be shortly dealt with (part 2) and then, in part 3, after the survey of syntactic functions of

¹ This paper was financially supported by the grant GAČR No. P406/10/0434.

infinitive and their frequency distribution, the conditions for the identification of the main syntactic functions will be formulated in several statements.

2. Identification of ambiguous infinitival forms

If syntactic functions of infinitival forms are to be studied in a sentence (or clause), these forms must be unambiguously identified as such, i.e. as infinitives. This is the task of an automatic morphological analysis and POS and morphological disambiguation. In contemporary Czech, only 44 infinitival forms of all verb lexemes are POS ambiguous (in the SYN2010 corpus, there are 19626 distinct verbal lemmas). These forms are ambiguous as follows:

- 42 with nominal forms, e.g. *stát* ‘V-stand, N-state’; *obrat* ‘V-bereave, N-turn/turning-point’; the whole list can be found in Petkevič 2006;
- 2 with numerals, e.g. *pět* and *pěti* ‘V-sing, Num-five’.

At present, the best tagging system of Czech is able:

(a) to disambiguate with 99,99% accuracy 37 ambiguous infinitival forms, e.g. *nemoci* ‘V-not_to_be_able, N-diseases’; *volat* ‘V-call, N-crop’; *pět/pěti* ‘V-sing, Num-five’;

(b) to disambiguate with at least 90% accuracy the following 7 ambiguous infinitival forms:

stát ‘V-stand, N-state’; *moci* ‘V-can, N-power’; *růst* ‘V-grow, N-growth’, *obrat* ‘V-bereave, N-turn/turning-point’; *drát* ‘V-strip/scramble, N-wire’; *vzrůst* ‘V-grow up, N-rise’; *srůst* ‘V-heal up, N-adhesion’.

The following example illustrates at least the complexity of the disambiguation of the verb/noun *stát*.

Example 1

(1)a. *Studenti budou hradit část ceny, kterou zaplatil stát_{Noun} za jejich vzdělání.*
‘Students will reimburse part of the cost which the state paid for their education.’

(1)b. *Oběti se můžeme stát_{Verb} snadno.*
‘Victim we can become easily.’

3. Syntactic functions of infinitive

In this main part of the study the survey of syntactic functions under investigation will be presented as well as a frequency distribution of these functions (part 3.1), and then analysis of necessary and sufficient conditions for the identification of syntactic functions will follow (part 3.2).

3.1 Repertory and frequency of syntactic functions of infinitive

If a verb form is unambiguously recognized and tagged as infinitive, it is possible to identify its syntactic function in a POS and morphologically annotated corpus. Without any theoretical discussion, I distinguish the following main syntactic functions of infinitive (in accordance with Šmilauer 1966): subject, nominal predicate, object, attribute, verbal complement and independent syntactic constituent, and one morphosyntactic function: part of periphrastic future.

First, the following frequency distribution of infinitival (morpho)syntactic functions was calculated on the basis of syntactic tagging of the working treebank SYNT and also manually verified in a sample of 500 POS and morphologically annotated sentences from the SYN2010 corpus.

Approximate frequency distribution of (morpho)syntactic functions of infinitive in %

Object:	69 %
Subject/NomPred:	13 %
Periphrastic future:	8 %
Independent synt. const.:	7 % – independent syntactic constituent, or the function cannot be determined
Attribute:	3 %
Verbal complement:	insignificant
Other cases:	insignificant

In the table, the subject and nominal predicate functions are not distinguished. The independent syntactic constituent includes also the cases where the function of infinitive cannot be distinguished within a clause (see restrictions below). In the sequel, the following main frequent (morpho)syntactic functions of active and passive² infinitive will be dealt with:

- infinitive as part of periphrastic future – special morphosyntactic function
- infinitive as subject / nominal predicate
- infinitive as attribute
- infinitive as object.

² Active and passive infinitive in Czech differ in form and function. The active infinitive is formed by one word form only (e.g. *pozvat* ‘to invite’), whereas the passive infinitive is formed by two word forms: infinitival form *být* ‘to be’ and a passive participle form of the autosemantic verb expressing number and gender (e.g. *být pozván* ‘to be invited’). Both active and passive infinitives can be assigned syntactic functions from the same repertory.

3.2 Analysis of syntactic functions of infinitive

The analysis will be restricted to syntactic functions of infinitives that can be identified within a clause, i.e. only the cases where the infinitive is not a syntactic head of a clause will be investigated. This means that there exists an infinitival syntactic governor in a clause which is to be identified.

It is assumed that the data on which the analysis will be performed has the following properties:

- sentences are lemmatized and POS and morphologically tagged;
- clause separators in sentences are annotated.

Definition 1. A *clause* in a sentence is a sequence of word forms between the left clause separator and the closest right clause separator. Clause separators can only be elements from the following set: (a) conjunction, (b) punctuation mark, (c) formal tag of the beginning/end of sentence.

Note 1. This notion of a clause is a deliberate simplification with respect to the classical term. Thus, in my considerably restricted view a clause cannot contain a clause separator, i.e. it cannot contain an embedded clause.

Example 2

(2) <S>_{ClauseSep} *Myslím* ,_{ClauseSep} *že sčítat škody budou několik dní* ._{ClauseSep} </S>
 ‘I think that to count damages they-will several days.’

In (2) there are two clauses:

- (a) *Myslím* which is enclosed in <s> and comma as its left and right clause separator, respectively;
- (b) *že sčítat škody budou několik dní* which is enclosed in comma and the full-stop as its left and right clause separator, respectively.

3.2.1 Infinitive as part of periphrastic future

In Czech, periphrastic future is formed by the pair:

- future form of the verb *být* ‘to be’ in 1st/2nd/3rd person singular/plural from the set VFutByt = {*budu/nebudu* ‘I shall be/I shall not be’, *budeš/nebudeš* ‘you will be/you will not be’, *bude/nebude* ‘(s)he/it will be/‘(s)he/it will not be’, *budeme/nebudeme* ‘we shall be/we shall not be’, *budete/nebudete* ‘you will be/you will not be’, *budou/nebudou* ‘they will be/they will not be’}
- active infinitive of an *autosemantic verb*.

Both word-order variants:

VFutByt Inf and *Inf VFutByt*

are possible:

(3)a. *My budeme_{VFutByt} nadále plnit_{InfPeriphFut} ty smlouvy a samozřejmě ...*

‘We shall furthermore comply with the contracts and of course ...’

(3)b. *Já myslím, že sčítat_{InfPeriphFut} škody budou_{VFutByt} několik dní.*

‘I think that to count the damages they-will several days.’

Thus, the infinitival form is part of a compound predicate periphrastically expressing future; here the infinitive has a morphosyntactic function. Necessary conditions under which the infinitival form can/cannot be part of periphrastic future will now be studied. The first statement about Czech is a well-known fact:

Statement 1. If an infinitive of verb V is *part of the form of periphrastic future*, then V is *not a perfective verb* (i.e. it may be an imperfective or biaspectual verb).

Example 3

(4) *Důležité_{NomPred} bude_{VFutByt} zachytit_{InfPerf} úvod jara.*

‘It will be important not to miss the beginning of the spring.’

As the infinitive *zachytit* is a form of the perfective verb in Czech, the construction *bude zachytit* cannot form periphrastic future; thus in (4) *zachytit* is subject, *bude* is a copula, *důležité* is a nominal predicate.

The following statement concerns negative infinitival forms:

Statement 2. If an infinitive is *negated*, then it is highly probable that it does not form periphrastic future with a future form of the verb *být* ‘to be’, i.e. from the VFutByt set.

Example 4

(5)a. *Nejlepší bude_{VFutByt} nepouštět_{InfNeg} se do řeči.*

‘The best will be not to strike up a conversation.’

(5)b. *Nejrozumnější bude_{VFutByt} nemluvit_{InfNeg} o zítřku.*

‘The most reasonable will be not to talk about tomorrow.’

In (5)a. and (5)b. both superlative forms *nejlepší* ‘best’ and *nejrozumnější* ‘the

most reasonable' are nominal predicates, respectively, *bude* is the copula and the infinitives *nepouštět* and *nemluvit* are subjects, respectively.

There is no strict grammatical rule in Czech stating that a negated infinitive cannot be part of periphrastic future, but it is almost always the future form from the VFutByt set that is negated, rather than the infinitival one. There exists, however, a very small group of verbs whose negated infinite is part of periphrastic future:

VInfNeg = {*nedostávat se* 'be missing', *nesnášet* 'detest', *nesouhlasit* 'disagree', *nelíbit se* 'dislike'}.

Example 5

(6)a. *zhoršení, protože se bude_{VFutByt} nedostávat_{InfNeg} 5,7 miliardy korun.*
'the deterioration, because will be missing 5.7 billion crowns.'

(6)b. *První se s nimi občas dostane do sporů, druhý je bude_{VFutByt} nesnášet_{InfNeg}.*
'The one will occasionally quarrel with them, the other will detest them.'

In (6)a. and (6)b. the pair *bude nedostávat* 'will be missing' and *bude nesnášet* 'will detest' form periphrastic future. Let us note that the verbs in the VInfNeg group do not form a coherent semantic class.

The Statement 2 can now be reformulated in a more precise way:

Statement 2a. If Inf is a negated infinitive not belonging to the VInfNeg set, then Inf and a future form of the verb *být* 'to be', i.e. from the VFutByt set, do not constitute periphrastic future.

3.2.2 Infinitive as subject / nominal predicate

First, necessary conditions for an infinitival form having the function of subject or nominal predicate are stated:

Statement 3a. If an infinitive has the function of *subject* in a clause, the finite predicate in the same clause is *in 3rd. pers. sg.* (+ *neuter* for Slavic l-type past participles that generally express gender and number).

Statement 3b. If an infinitive has the function of *subject* in a clause, no syntactic noun *Nom* can, of course, be subject of the clause. In this case, a nominal element *Nom* in the 3rd. pers. nominative singular can be only:

- part of a comparative construction (*jako Nom / než Nom* 'like Nom / than Nom')
- part of a parenthesis
- nominative of nomination.

Example 6

(7) *Policisté zatím neupřesnili, jak se řidiči_{NounDat} podařilo_{FinVerbSg} vyvážnout_{InfSubj}.*
 ‘The police have not yet specified how to the driver it succeeded to escape.’
 ‘The police have not yet specified how the driver succeeded in escaping.’

In (7) *podařilo* ‘succeed’ is a finite verb in singular, the form *řidiči* ‘driver’ is in dative singular (morphologically, the form itself can also express locative singular and nominative/vocative/instrumental plural, but here the form is not in nominative plural, because in Czech the finite verb and its subject must agree in number) and the infinitive *vyvážnout* is subject. The set of predicates enabling the subject to be formed by an infinitive is, however, considerably restricted:

Statement 4. If *subject* is formed by an infinitive in a clause, then the finite predicate belongs to a very small set of verbs of the two types A and B:

A. Verbs that do not require reflexive particle *se* expressing deagentivity, e.g. *být/bývat* ‘be’; modal verbs: *jít* ‘go’, *lze* ‘can be’; phase verbs: *začít/začínat* ‘begin, start’, *přestat/přestávat* ‘stop, cease’; other verbs: *znamenat* ‘mean’; *bavit* ‘be interested’, *mít smysl/význam/cenu* ‘make sense’ etc.

Inherent reflexive verbs:

zdát se ‘seem’, *dařit se/podařit se/zdařit se* ‘succeed’, *líbit se/zalíbit se* ‘like’, *hnusit se* ‘detest’, *zamlouvat se* ‘like/taste’, *poštětit se* ‘be lucky’, *dát se* ‘be possible’, *zachtít se* ‘want’ etc.

B. Verbs that require reflexive particle *se* expressing deagentivity, e.g.:

modal verbs: *moci* ‘can’, *muset* ‘must’, *mít* ‘to be to’, *chtít* ‘want’, *smět* ‘may’; other verbs: *ukázat* ‘manifest itself’, *rozhodnout* ‘decide’, *vyplatit* ‘pay’, *doporučovat* ‘recommend’ etc.

The complete list of such verbs has been developed by the author.

The validity of Statement 4 is illustrated by the following sentences where each infinitive has the syntactic function of subject and finite verbs belong to the A or B set specified above.

Example 7

(8) *Odejít_{InfSubj} pro něho znamenalo_{FinVerb} porážku.*
 ‘To leave for him meant defeat.’

(9) *Stále neúplný oltář se podařilo_{FinVerb} kněžím včas ukryt_{InfSubj}.*

‘A still incomplete altar it-succeeded to the priests to hide in time.’
 ‘The priests succeeded in hiding the still incomplete altar in time.’

(10) *Nám se líbilo/zamlouvalo*_{FinVerb} *si hrát*_{InfSubj}.

‘To us it-liked to play.’

‘We liked to play.’

(11) *Mělo*_{FinVerb} *se kvalifikovaně rozhodnout*_{InfSubj}.

‘It should have been in a qualified way to decide.’

‘It should have been decided in a qualified way.’

(12) *Chce*_{FinVerb} *se mi zvracet*_{InfSubj}.

‘It wants me to vomit.’

The verb *být* ‘be’ as a finite predicate whose subject is formed by an infinitive will now be discussed in more detail. Infinitive has the function of subject (possibly nominal predicate) in the following three patterns/constructions with the verb *být* (Pattern 1, 2, 3 below):

Pattern 1: *je*_{Vby3sg} *AdvPart** *nutné/zakázáno* *AdvPart** *pracovat*_{Inf}

‘is *AdvPart** necessary/forbidden *AdvPart** to work’

The word-order of elements in the pattern is free, virtually no grammatical restrictions apply here. Generally, the pattern is as follows:

Vby3sg *AdvPart** (AdjModEval | AdvPred | VPasSgNeut) *AdvPart** Inf

where:

(i) Vby3sg is a finite verbal form of the lexeme *být* ‘be’ in 3rd pers. sg. (+ neuter for past participles of the l-type) belonging to the following set:

{*je/není* ‘is/is not’, *bude/nebude* ‘will be/will not be’, *bylo/nebylo* ‘was/was not’, *bývalo/nebývalo* ‘used to be/used not to be’, *bývá/nebývá* ‘is/is not’}

(ii) the adjective AdjModEval is modal or evaluative and it must be in nominative sg. neut. It is a member of the following set of forms:

{*možné* ‘possible’, *nutné* ‘necessary’, *nutno* ‘necessary’, *těžké* ‘difficult’, *vhodné* ‘appropriate’, *zajímavé* ‘interesting’, *dobré* ‘good’ etc.}

(iii) AdvPred is a predicative adverb from the set:

{*třeba/potřeba/zapotřebí* ‘necessary’, *namístě* ‘appropriate’, *těžko/zatěžko* ‘difficult’ etc.}

(iv) VPasSgNeut is a passive participle in sg. neuter from a small set:

{*zakázáno* ‘forbidden’, *umožněno* ‘enabled’, *dovoleno* ‘permitted’, *souzeno* ‘doomed’, *nařízeno* ‘recommended’ etc.}

- (v) AdvPart* is a (possibly empty) sequence of adverbs and/or particles.
 (vi) Inf is an infinitival form.

With Pattern 1 having been specified the following statement can be formulated.

Statement 5. If there is Pattern 1 in a clause, Inf is *subject*.

Several illustrative examples follow:

Example 8

(13) *Objekty je_{Vby3sg} nutné_{AdjModEval} dobře zrekonstruovat_{InfSubj}.*
 ‘The objects it is necessary to rebuild.’

(14) *Je_{Vby3sg} nám zakázáno_{VPasSgNeut} okamžitě odejít_{InfSubj}.*
 ‘It is to us ordered to leave immediately.’

It is important to realize that with passive forms in neuter singular not belonging to the VPasSgNeut set such as *rozhodnuto* the Statement 5 cannot be applied unambiguously:

(15)a. *Bylo rozhodnuto demonstrativně odejít_{InfSubj}.*
 ‘It was decided to leave in a demonstrative way.’

(15)b. *Grémium_{Subj} bylo rozhodnuto nezúčastnit_{InfObj} se jednání.*
 ‘The committee was resolved not to participate in negotiations.’

Whereas in (15)a. the infinitival form *odejít* is a subject as indicated, in (15)b. the infinitival form *nezúčastnit se* is an object and *grémium* is a subject.

Pattern 2:

naším cílem/náš cíl AdvPart je/bylo/bude AdvPart* usnadnit...*
 ‘our objective AdvPart* is/was/will be AdvPart* to facilitate...’

The word-order of elements in the pattern is free again with virtually no grammatical restrictions applying here. Generally, the pattern is as follows:

NInstr/NNom AdvPart* Vby3sg AdvPart* Inf

where:

- (i) NInstr/NNom is a syntactic noun in instrumental/nominative case;
 (ii) Vby3sg is a finite verbal form of the lexeme *být* in 3rd pers. sg. (+ neuter for past participles of the l-type);

- (iii) Inf is an infinitival form but not of the verb of perception from the set: {*poznat* ‘recognize’, *slyšet* ‘hear’, *vidět* ‘see’, *znát* ‘know’}
- (iv) AdvPart* is a (possibly empty) sequence of adverbs and/or particles.

Pattern 2 allows us to specify the following statement:

Statement 6. If there is a Pattern 2 in a clause, then:

- a. if NInstr, then Inf has the function of *subject*;
- b. if NNom, then Inf has the function of *nominal predicate*.

Example 9

(16) *Naší snahou*_{NInstr} *zatím bude*_{Vby3sg} *získávat*_{InfSubj} *informace*.
 ‘Our endeavour for the time being will be to gather information.’

In addition to the specified interpretation of the function of the infinitive *získávat* in sentence (16) there is yet another interpretation where the subject in singular is not lexically expressed and the pair *bude získávat* is periphrastic future rather than subject:

(16)a. *Naší snahou*_{NInstr} *zatím bude*_{Vby3sg} *získávat*_{InfPeriphFut} *informace*.
 ‘By our endeavour for the time being he/she/it will be gathering information.’

Thus, sentence (16) comprising Pattern 2 is syntactically ambiguous.

Pattern 3. Let us define the small set of infinitives of the verbs of perception:
 VPerInf = {*poznat* ‘recognize’, *slyšet* ‘hear’, *vidět* ‘see’, *znát* ‘know’}

and let us have the pattern:

Vby3sg AdvPart* Vpercinf (Clause | Noun_{Nom|Acc})

where:

- (i) Vby3sg is a finite verbal form of the lexeme *být* in 3rd pers. sg. (+ neuter for past participles of the I-type);
- (ii) Vpercinf is an element of the VPerInf set
- (iii) Noun_{Nom|Acc} is a noun in the nominative or accusative case
- (iv) AdvPart* is a (possibly empty) sequence of adverbs and/or particles.

The forms from the VPerInf set allow for the following constructions:

(17a) *Je vidět*_{NomPred} *Sněžka/Říp*_{NomSubj}.

(17b) *Je vidět_{Subj} Sněžku/Říp_{AccObj}.*
 ‘It is possible to see Sněžka/Říp.’

There exist two interpretations of the morphologically ambiguous word *Říp*: nominative or accusative singular; on the level of syntax they constitute subject and object, respectively.

Moreover, the verbs from the VPerInf set allow for two plausible interpretations also in case of a dependent clause (which lacks case) taking up the position of the noun (like *Říp* above):

(18a) *Na Tomášovi s Jirkou bylo_{Copula} vidět_{InfSubj}, (že se teprve sehrávají)_{Obj}.*

(18b) *Na Tomášovi s Jirkou bylo_{Copula} vidět_{NomPred}, (že se teprve sehrávají)_{Subj}.*

‘On Tom and George it was to see, that they are only getting used to playing together.’

In (18a) and (18b) the infinitive *vidět* and the clause following it have different syntactic functions as indicated.

Now the statement concerning Pattern 3 can be presented:

Statement 7. If there is a pattern of Pattern 3 in a clause, then:

- a) if $\text{Noun}_{\text{Nom|Acc}}$ is in the nominative case, then Vperinf has the function of nominal predicate (see (17a));
- b) if $\text{Noun}_{\text{Nom|Acc}}$ is in the accusative case, then Vperinf has the function of subject (see (17b));
- c) if $\text{Noun}_{\text{Nom|Acc}}$ is nominative/accusative ambiguous, then Vperinf is syntactically ambiguous: it has the function of nominal predicate or subject (see (17a) and (17b));
- d) if there is a Clause instead of $\text{Noun}_{\text{Nom|Acc}}$, then Vperinf is ambiguous as in alternative (c) above: it has the function of the nominal predicate or subject (see (18a) and (18b)).

3.2.3 Infinitive as attribute

In approximately 3% of cases an infinitive can have the function of attribute. The set of nouns that can be modified by an attributive infinitive comprises only tens of action/deverbal nouns belonging to the set NAttrInf:

NAttrInf = {(ne)možnost ‘(im)possibility’, potřeba ‘necessity’, úkol ‘task’, snaha ‘endeavour’, šance ‘hope’, právo ‘right’, cíl ‘objective’, příležitost ‘duty’, povinnost ‘occasion’ etc.}

Moreover, such an attributive infinitive almost always follows its governor from NAttrInf in the same clause; only in very marked cases it can precede it. Both

the governing noun and the infinitive can be modified, i.e. they need not take up an adjacent position in a clause:

NAttrInf *NAttrInfDep* *InfDep* Inf

Here *NAttrInfDep* is a set of constituents depending on NAttrInf, *InfDep* is a set of constituents depending on Inf.

The following probabilistic statement specifies a condition under which an infinitive has not the attributive function:

Statement 8. If in a clause an infinitive Inf is *not preceded* by a lemma belonging to NAttrInf, then Inf is very rarely an attribute.

However, it is difficult to specify a general sufficient condition for Inf to be an attribute of a noun in a clause because nouns from NAttrInf differ in their ability to attach the attributive infinitive. We must be satisfied with the following probabilistic statement:

Statement 9. If in a clause there is an adjacent pair: *Noun Inf* where *Noun* belongs to *NAttrInf* and *Inf* is the leftmost verb in the clause, it is highly probable that *Inf* is an attribute of *Noun*.

Example 10

(19) *Úkol*_{Noun} *vybrat*_{InfAttr} *tu nejkrásnější čeká na porotu.*
 ‘The task to elect the most beautiful one waits for the jury.’

In (19) the infinitival form *vybrat* immediately follows the noun *úkol* belonging to the *NAttrInf* set and modifies it as its attribute.

After the specification of several necessary concepts we can now specify yet another statement concerning the periphrastic future function of an infinitive. This time the set of sufficient conditions for an infinitive to co-form periphrastic future is presented:

Statement 10. Let all the following conditions hold in a clause:

- (a) Inf is an infinitival form of a imperfective verb;
- (a) NAttrInf is a set of lemmas of action/deverbal nouns that can be modified by an attributive infinitive (see above);
- (b) The future form VFutByt of the verb *být* is in plural:

{forms *budeme/nebudeme* ‘we shall be/we shall not be’, *budete/nebudete* ‘you will be/you will not be’, *budou/nebudou* ‘they will be/they will not be’}

or

in the 1st or 2nd person singular: {forms *budu/nebudu* ‘I shall be/I shall not be’, *budeš/nebudeš* ‘you will be/you will not be’}

(c) in the word-order sequence VFutByt ... Inf within the clause there is no Noun from the NAttrInf set: thus VFutByt ... ~~NAttrInf~~ ... Inf

(d) if in the clause there is a word-order sequence Inf ... VFutByt, there is no Noun from the NAttrInf set standing in front of Inf in the sentence:

<SentStart> ... ~~NAttrInf~~ ... Inf ... VFutByt

(e) in the clause, there is no element from the following sets: AdjModEval, AdvPred, VPasSgNeut (cf. sect. 3.2.2).

Then the form VFutByt co-forms periphrastic future with the closest infinitival form Inf.

Example 11

(20) *Iráčané si budou_{VFutByt} opravdu muset_{InfPeriphFut} vládnout sami.*
‘the Iraqis will have to rule themselves’

In (20) the pair *budou ... muset* forms periphrastic future of the word *muset*.

But (!) in sentence:

(21) *na něco budou_{VFutByt} možnosti_{NAttrInf} čerpat_{InfAttr} peníze*
‘for something, however, there will be possibilities to use money’

the pair *budou ... čerpat* does not form periphrastic future of the verb *čerpat*, because the intervening noun *možnosti* belonging the NAttrInf set blocks this possibility: the infinitive *čerpat* modifies the noun *možnosti* as its attribute.

If the future form is in 3rd person singular (*bude* ‘it will be’), the situation is more complex because the infinitive can be subject or nominal predicate, the form *bude* being the copula, cf.:

(22) *Nejlepší bude_{Copula} na zápas rychle zapomenout_{InfSubj}.*
‘Best of all will be to forget the match as soon as possible.’

In (22) the perfective verb *zapomenout* ‘forget’ can never be part of a periphrastic future; here it is *subject*.

The difficult problem arises, however, with imperfective verbs:

(23) *Řekla, že nejlepší bude_{Copula} držet_{InfSubj} se svého...*
 ‘She said that best will be to stick to one’s own...’

In (23) *nejlepší* is a nominal predicate, *bude* is a copula and the infinitive *držet* is the subject of the clause introduced by the conjunction *že*. Similarly as in sentence (16), a syntactic ambiguity arises here: *nejlepší* can also be regarded as subject, the pair *bude držet* as periphrastic future.

3.2.4 Infinitive as object

The object function is the most frequent function of infinitive (cca 70 %, cf. sect. 3.1 above). An infinitival form having this function modifies verbs from a special class and also adjectives derived from these verbs. These verbs (and derived adjectives) must have an object valency that can be realized by an infinitive. For Czech, the whole set of such verbs has been specified and this set is the pivotal element of the following statement:

Statement 11. If in a clause there is:

- a) a finite verb in plural or in the 1st or 2nd person belonging the set of verbs VInfVal whose object can be realized by an infinitive, where:
 $VInfVal = \{moci \text{ ‘can’}, muset \text{ ‘must’}, mít \text{ ‘have’}, chtít \text{ ‘want’}, dokázat \text{ ‘be able’}, snažit se \text{ ‘endeavour’}, hodlat \text{ ‘intend’}, rozhodnout se \text{ ‘decide’ etc.}\}$
- b) an infinitive Inf
- c) no noun from the NAttrInf set standing in front of Inf in the clause,

then Inf has the function of *object* in both word-order variants:

VInfVal ... InfObj and InfObj ... VInfVal

Example 12

(24) *že jsme se snažili_{VInfVal} je informovat_{InfObj}.*
 ‘that we tried to inform them’

In (24) the verb *snažili* is a plural form of the verb *snažit se* belonging to the VInfVal set and *informovat* is its object because the conditions of the Statement 11 are met.

In addition to the verbs from the VInfVal set, several passive verbal forms and adjectives can also have infinitives as their objects:

Statement 12. If in a clause there are the following elements:

- (a) a passive verbal form from the set VPassObjInf which is not in neuter singular

or

an adjective belonging to the set of adjectival lemmas AdjObjInf not in neuter singular, where

VPassObjInf = {(do)nucen/a/i/y ‘compelled’, připraven/a/i/y ‘ready’, oprávněn/a/i/y ‘authorized’, rozhodnut/a/i/y ‘resolved’, etc.}

AdjObjInf = {schopný/schopen ‘able’, ochotný/ochoten ‘willing’, povinen ‘accountable’, zvyklý ‘accustomed’ etc.};

(b) an infinitival form Inf following an element from AdjObjInf or VPassObjInf;

and, moreover, there is no noun from the NAttrInf set standing in front of Inf in the clause,

then Inf has the function of *object*.

Example 13

(25) *Armáda je připravena_{VPassObjInf} posílit_{InfObj}.*
‘The army is ready to strengthen.’

The passive form *připravena* of the verbal lexeme *připravit* is modified by the infinitival form *posílit* in the object function.

4. Conclusion

In this paper the main (morpho)syntactic functions (periphrastic future, subject/nominal predicate, attribute and object) of infinitive were investigated. The results were formulated in the form of statements containing necessary and sufficient conditions for the identification of these functions. The statements are to be implemented as formal rules within a Czech formal grammar underlying syntactic analysis of Czech used for the development of a Czech treebank.

References

- Novák, Václav & Zdeněk Žabokrtský. 2007. Feature Engineering in Maximum Spanning Tree Dependency Parser. In *Proceedings of the 10th International Conference on Text, Speech and Dialogue*. Plzeň: Západočeská univerzita. Berlin, Heidelberg: Springer-Verlag, LNCS 4629, 92–98.
- Petkevič, Vladimír. 2006. Automatické rozpoznání infinitivu. Případová studie jako příspěvek k automatické disambiguaci českých textů. In F. Čermák Fr. & R. Blatná (eds.): *Studie z korpusové lingvistiky I. Korpusová lingvistika: Stav a modelové přístupy*. Praha: Nakladatelství Lidové noviny, 226–253.
- Šmilauer, Vladimír. 1966. *Novočeská skladba*. Praha, SPN.