DEPARTMENT OF COMPUTATIONAL LINGUISTICS

INSTITUTE FOR BULGARIAN LANGUAGE "PROF. LIYBOMIR ANDREYCHIN"

BULGARIAN ACADEMY OF SCIENCES

Author's abstract of the dissertation for the award of the Educational and Scientific Degree "PhD"

BULGARIAN-ENGLISH GRAMMATICAL PARALLELS IN RELATION TO MACHINE TRANSLATION

Professional field
2.1. Philology

Scientific field
General and Comparative Linguistics
(Computational Linguistics)

Author: Todor Lazarov
Scientific supervisor: prof. Svetla Koeva, PhD

Members of the scientific committee:
prof. Maksim Stamenov, PhD
prof. Yovka Tisheva, PhD
prof. Krasimira Aleksova, PhD
prof. Radka Vlahova, PhD
ass. prof. Ekaterina Tarpomanova, PhD

Reserve Members of the Scientific Committee:
prof. Velka Popova, PhD
prof. Diana Blagoeva, PhD

The dissertation has been discussed on an extended session of the Department of Computational Linguistics of the Institute for Bulgarian Language "Prof. L. Andreichin" at the Bulgarian Academy of Sciences on 27.01.2020 and it was directed for public defence by Protocol No 2 / 30.01.2020 of the Scientific Council of the Institute for Bulgarian Language" Prof. L. Andreychin "at the Bulgarian Academy of Sciences.

The dissertation consists of 197 pages. The dissertation consists of Introduction, Four Chapters, Conclusion, Bibliography and annex.

Contents

I. General Characteristics of the Dissertation

I.1. Relevance and Motivation for the Research

Digital processing of linguistic data has become increasingly popular in recent years and is being used more widely not only in the field of linguistics but also in a number of related fields. One of the reasons for the growing popularity of digital processing is that it allows researchers to process large amounts of textual data and identify different textual features, common features, patterns that may be difficult or even impossible to identify by the methods of classical analysis. Increasingly popular and similar trend is the creation of different language corpora that contain different types of language data for the purpose of natural language processing and creation of applicable language models.

The increasing popularity of machine translation in the recent decades has been driven by the need for quick access to different types of information in different languages, volume and types. The main advantages of machine translation systems are consistency of translation, improved productivity and reduced costs. A major disadvantage of machine translation systems is that they have been developed and are available mainly for the more popular languages such as English, French, German, Spanish, Russian, Arabic and others.

The purpose of the dissertation is related to the research of the Bulgarian-English grammatical parallels in relation to machine translation. The scope of the research is limited to the temporal systems in both languages: the means of expressing temporal relations in Bulgarian and English and the means of transfer of the grammatical meaning of the temporal forms from Bulgarian to English. On the one hand, the research in relation to the approaches and methods used in machine translation is motivated by the possibilities that the scientific field offers for the systematic and formal analysis of complex linguistic phenomena, such as the temporal systems in both languages. On the other hand, the insufficient availability of adequate parallel language resources that contain information for translational correlation of verb forms between Bulgarian and English suggests the theoretical and applied aspect of the study.

I.2. Research purposes

The main purpose of the dissertation is to present a unified theoretical description of the temporal systems in Bulgarian and English and to examine the theoretically derived hypotheses for translation equivalents by statistical analysis of data from a parallel corpus with aligned temporal forms between Bulgarian and English: "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

The results of the research can be applied in improving the operation of computer aided translation systems, as well as in the selection of monolingual and parallel training data for machine translation systems.

The following terminological clarifications are introduced for the purposes of the dissertation: the term grammatical tense denotes the traditional nine-member tense system in Bulgarian and the sixteen-member tense system in English (and infinitive). The nine tenses in Bulgarian are: Praesens (сегашно време), Aorist (минало свършено време), Imperfect (минало свършено време), Perfect (минало неопределено време), Plusquamperfectum (минало предварително време), Futurum (бъдеще време), Futurum Exactum (бъдеще предварително време), Futurum Praeteriti (бъдеще време в миналото) and Futurum Exactum Praeteriti (бъдеще предварително време в миналото) and the sixteen English tenses are: Present Simple Tense, Present Continuous Tense, Present Perfect Tense, Present Perfect Continuous Tense, Past Simple Tense, Past Continuous Tense, Past Perfect Tense, Past Perfect Continuous Tense, Future Simple Tense, Future Continuous Tense, Future Perfect Tense, Future Perfect Continuous Tense, Future Simple Tense in the Past, Future Continuous Tense in the Past, Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past.

The following theoretical view is adopted: the grammatical tense (also hypercategory tense) expresses different (temporal) relations and it can be considered as a set of three morphological categories in Bulgarian: tense, type of action and taxis, and two morphological categories in English: tense and aspect. The **morphological tense** (also proper tense) in **Bulgarian** "is based on the relation between an action and utterance" (Kutsarov 2007: 244), this relation can express "simultaneity or a-synchronicity (non-simultaneity)". The **morphological tense in English** "expresses simultaneity or a-synchronicity of an action in relation to the act of speech" (Kabakchiev 2003: 132). The morphological category **"type of action"** (resultativity) **in Bulgarian** expresses "a relation between an event and a result from an event at the same orientation moment" (Kutsarov 2007: 249) that can be present, past, future

or future in the past. The morphological category **aspect in English** expresses a relation between "the continuity and the non-continuity of an event [...] in relation to a giver point of reference" (Valeika and Buitkiene 2003: 84) in the Present, the Past, the Future or the Future in the Past. It is necessary to specify that the category type of action in Bulgarian and the category aspect in English represent two different categories with different semantics and categorical structure. The morphological category of **taxis in Bulgarian** expresses "the relations between actions and results from actions that are oriented to an orientation moment in the past (Past Tense or other lexical equivalent) and actions that are not orientated to an orientation moment in the past" (Kutsarov 2007: 255).

The presented terminological definitions are perceived as most clearly defining the meaning of the respective morphological categories for Bulgarian and English. As none of the cited authors provides definitions regarding the comparison of temporal systems in Bulgarian and English, the generalizations cannot be considered as typological.

The term **temporal relations** denotes the relations expressed by the three morphological categories: tense, type of action and taxis in Bulgarian, and the morphological categories: tense and aspect in English. The term **temporal system** denotes the system consisting of the morphological categories tense, type of action and taxis in Bulgarian and tense and aspect in English. The term **temporal forms** denotes the verb forms in Bulgarian and English that are part of the respective temporal systems.

The main purpose of the dissertation is achieved through the goals and objectives described in the individual chapters of the dissertation.

The purpose in Chapter Two "Bulgarian-English Grammatical Parallels in Regard to the grammatical tense" is to define an unified conception for the study of the temporal forms in both languages in order to represent a formal semantic description of the temporal meanings in the languages in contrastive plan and to develop a hypothesis for the possible semantic equivalents. The formal semantic description of the temporal meanings in both languages will be used to construct a hypothesis for the Bulgarian-English grammatical parallels at the level of the temporal forms.

The specific tasks for achieving the purpose of the second chapter of the thesis are:

1. To study and compare the different theoretical descriptions of the morphological categories tense, type of action and taxis in Bulgarian and tense and aspect in English.

2.   To develop a unified conception for description of the temporal systems in Bulgarian and English, the morphological categories that construct them and the relations between their members.

3.   To develop a formal semantical description of the temporal meanings in both languages.

4.   To outline the grammatical parallels between the formal-semantic descriptions of temporal meanings in Bulgarian and English, on the basis of which a hypothesis for the translation equivalents will be created and for the semantic transfer of grammatical information during translation.

The purpose of the research in Chapter Three, "Creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is to create a linguistically annotated parallel corpus containing information for translation of the Bulgarian temporal forms with their English equivalents. The following tasks must be completed in order to accomplish the purpose:

1.   Analysis of the main features of machine translation and the requirements for language resources used in machine translation systems.

2.   Review of existing parallel resources for Bulgarian and English in order to select appropriate resources for the purposes of the research.

3.   Developing criteria for creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

4.   Developing annotation conventions for supplementing the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" with linguistic information on the aligned verb forms.

5.   Practical work for creation and annotation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

6.   Description of the main characteristics of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

The main purpose of the research in Chapter Four is to present a statistically based analysis on the temporal forms in Bulgarian and English individually and in contrastive plan. The analysis is: in respect to the data for the Bulgarian and English temporal forms separately; in respect to the transfer of information about morphological categories tense, type of action and taxis in Bulgarian to the English morphological categories tense and aspect, as well as in

respect to transfer of information about the morphological categories person, number and gender and the lexical-grammatical category type of the verb in during translation from Bulgarian to English. The specific tasks for achieving the purpose of the fourth chapter of the thesis are:

1.   Data extraction and statistical analysis of the Bulgarian temporal forms, divided by temporal categories and thematic areas in the Corpus.

2.   Data extraction and statistical analysis of the English temporal forms, divided by temporal categories and thematic areas in the Corpus.

3.   Statistically based analysis of the translation of temporal forms from Bulgarian to English.

4.   Forming a conclusion about the Bulgarian-English grammatical parallels with respect to the temporal forms.

5.   Presenting the main results of the analysis of the data in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" in the context of the theoretical description of the temporal systems in Bulgarian and English.

Another main purpose of the research is the practical application of the Corpus data in creation of translation models enriched with linguistic information for Bulgarian-English correlations between the temporal forms. In particular, the practical application of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" in a computer aided translation system. For achieving this purpose the following tasks presented in Chapter Five "Practical Application of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" are developed:

1.   Creation of terminological database that contains data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

2.   Conducting an experiment with the terminological database in a computer aided translation system.

3.   Evaluation of the results from the experiment.

In conclusion, the purpose of the dissertation can be presented as a set of tasks in relation to several conceptually separate parts. On the one hand, defining a unified concept for description of the temporal systems and formation of the verb forms in Bulgarian and English is a necessary prerequisite for comparing the two languages at theoretical level. On the other

hand, the development of a hypothesis, based on the theoretical description of temporal systems in both languages, provides the possibility to evaluate the extracted statistical linguistic data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

I.3.   Area and object of the research

The area of the research are the contemporary temporal systems in Bulgarian and English, the means of expressing the hypercategory of tense in both languages, the semantic transfer of the meanings of the morphological categories tense, type of action and taxis in Bulgarian to the morphological categories tense and aspect in English. The object of the research are the specific grammatical parallels between the temporal relations in Bulgarian and English, derived from the data in a parallel corpus designed for the purposes of the research.

The purpose of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is to present full and adequate linguistic information that can be applied for the creation of statistical translation model of the verb forms. The clarification should be made that the metadata in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" does not reflect the morphological categories evidentiallity, mood, voice positive or negative status. The choice of the particular direction of translation (from Bulgarian to English) is dictated by the complexity of Bulgarian verb forms compared to English.

I.4.   Research Method

In the recent years there is a tendency for a formal semantic description of language characteristics and for practical application of this description. The study of the hypercategory of tense in both languages provides an opportunity for a parallel formal semantic analysis of the temporal categories.

It should be noted that the formal study of the characteristics of the word formation in Bulgarian and English (individually and comparatively) is not new; there are researches focusing on more or less general or specific problems; there are attempts to present formal grammatical parallels for translation between Bulgarian and English, in respect to other parts of speech and in respect to verbs (Atanasova et al. 1963, Stambolieva 2015). In this dissertation the problem of the Bulgarian-English grammatical parallels between the temporal systems in both languages is studied through the methods of formal semantic analysis, corpus analysis and statistical modelling.

The chosen methods for analysis of the Bulgarian-English grammatical parallels at the level of temporal systems are closely related to the field of computational linguistics and to machine translation in particular. The research is influenced by modern trends in systematic analysis of linguistic phenomena through application of computational linguistics and statistical modelling (Paskaleva 1987: 32). The relevance of statistical modelling implies an integration of current trends in language studies and classical theoretical studies.

The grammatical parallels between Bulgarian and English and the formal semantic description of the differences and similarities of the temporal systems in the two languages are empirically verified on the basis of statistic data derived from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms". For this purpose the corpus analysis is combined with statistical modelling.

Machine translation relies on representative parallel corpora of sufficient volume to represent linguistic phenomena in their various manifestations. In order to create a statistical translation model, it is necessary to know what is the probability that a language unit (word, phrase) in the target language will be a translation of a language unit from the source language – p(e|f). In order to achieve this, it is necessary to develop algorithms that process information and language resources to extract linguistic data. As it has been stated, "corpora [...] represent a reliable source for observation, analysis and conclusions (supported by objective quantitative and distributional data) for [...] different linguistic phenomena (for their use in different styles, genres and thematic areas), as well as for automatic retrieval of language data, language relations and models" (Koeva 2014: 49).

The grammatical parallels between Bulgarian and English and the formal description of the differences and similarities of the temporal systems in the two languages is based on analysis of the data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

The results are experimentally verified in a computer aided translation software.

II.   Bulgarian-English parallels in respect to grammatical tense

Chapter Two of the dissertation presents a theoretical description of the temporal systems in Bulgarian and English and the formal-semantic characteristics of the two systems.

II.1. Tense as verbal category - conceptions and discussions

The description of the phenomenon of time has been a subject of discussion in both philosophy and linguistics. The Aristotelian tradition represents time as "a continuous line that flows from left to right, i.e. from the past to the future "(Manning et al. 2013: 233). The tense in linguistics is seen as a grammatical category that "grammatically represents the course of events or states through time by localizing possible events" (Comrie 1985: 9).

In both Bulgarian and English language studies, the category of tense has always been debatable in terms of its scope: inventory of grammatical meanings and encoding of grammatical information. Modern linguistic studies adopt the concept of so-called "hypercategories", consisting of interrelated categories with their own grammatical paradigms, which are cumulative in terms of form formation (Tsergova 2014: 195). The category of tense is identified with the semantic continuum "relation of the event and at the moment of speaking" (Tsergova 2014: 203), but within this meaning can be distinguished references to the direct and indirect relation of the event in regard to the moment of speaking, that can form two subcategories in temporality: verb tense and taxis (Tsergova 2012: 18-19).

Increasingly productive from a functional point of view are the theories that examine the verb categories as "complex functional-semantic structures with diverse linguistic instrumentarium" (Veiga 1991: 56) or hypercategories in which "between the finite elements of the relation expressed of one simple category one or more intermediary elements are possible" (Gerdzhikov 2000). "The complex categories (hypercategories) consist of syncategories that do not have equal status in the system due to the elements that construct their meanings (only finite elements, finite elements and intermediary elements, etc.)" (Aleksova 2003).

Regarding the theory of temporal logic, there are three possible temporal relations between the moment of the event to the moment of speaking (point of speech): "before the point of speech", "simultaneous with the point of speech" and "after the point of speech" (Reichenbach 1947: 287). Grammatical tense in a very strict sense contains only two semantic values: simultaneity and non-simultaneity (Reichenbach 1947: 291). Other temporal relationships are formed by additional categories such as aspect or relativity. The distinction between aspect and grammatical tense is not uniformly realized in different languages,

although they can be regarded as interdependent and interrelated (Hornstein 1990: 9). The aspect in Bulgarian is a lexical-grammatical category and formally remains outside the hypercategory of tense, while in English it is expressed by the morphological category of aspect (perfect and continuous tenses), and in some cases at lexical level (by phraseological verbs).

For the means of formal representation of H. Reichenbach's temporal system, the abbreviation S will be used to indicate the moment (time) of speaking (speech). The moment of the event expressed by the verb will be denoted by E. There are two possible relations between S and E: the relation of a sequence with two subdivisions –E precedes S and vice versa; and relation of simultaneity – E and S occur at the same time and S is included in E.

An important feature of H. Reichenbach's temporal system is that it is characterized by three temporal points: the previously defined points E and S and the so-called "point of reference" (R) (Reichenbach 1947: 293). In addition to the three temporal points (S, E, and R) the system includes two types of relations between them (simultaneity and succession). An important feature is that these types of relations can only be between points R and S or R and E, the relation between S and E is indirect and is built on other relations in the system (Klein 1994: 54). Therefore the point of reference R can be considered as a connecting node in the system. The relations between S and R, on the one hand, and E and R, on the other, according to the system of H. Reichenbach (Reichenbach 1947: 295) are presented in Table 1:

Relation	Meaning	Name		Relation	Meaning	Name
S, R	S is simultaneous with R	Present		E, R	E is simultaneous with R	Simultaneity
S – R	S precedes R	Future		E – R	E precedes R	Anteriority
R – S	R precedes S	Past		R – E	R precedes E	Posteriority

Table 1. *Description of the relations between the elements in the system of H. Reichenbach*

II.2. Characteristics of the Bulgarian verbs

Compared to other Slavic languages only in Bulgarian the verb retains significant part of the most important features of the Old Bulgarian language, but also undergoes further development (Nitsolova 2008: 224). According to the classification based on the morphological characteristics, the verb is a variable part of speech and a large number of word forms with multiple grammatical meanings are included in the verb paradigm. According to its syntactic function and characteristics, the verb is an independent part of speech, since it is the functional part of the sentence that expresses the relation between the other components of the sentence – it relates the arguments within the sentence (Nitsolova 2008: 224).

The main lexical division of the verbs is based on their functional lexical-grammatical characteristics and divides the verbs into main verbs and auxiliary verbs (Kutsarov 2007: 108). The verbs can be transitive or intransitive according to the possibility of a direct complement, if the action directly affects an object. According to expressing the grammatical meaning "person", the verbs are divided into personal and impersonal (Kutsarov 2007: 105).

The grammatical category type of the verb "allows one action to be presented in two ways: as a completed action and as an unfinished process. Therefore, the type of the verb is a morphological category that expresses the way in which the verb action proceeds in relation to the completeness (exhaustion, completeness, complexity)" (Garavalova 2003: 186). This peculiarity of the Bulgarian verbs has been a subject of different scientific conceptions and discussions. It is agreed that the division of the verbs in the Bulgarian language into perfective type and imperfective type is within the lexical-grammatical category type of the verb (Kutsarov 2007: 551). The division in this category divides the verbs according to whether or not the expressed action has a limit to limited verbs and unlimited verbs. "Only the limited verbs can be perfective and imperfective, the unlimited verbs are only imperfective" (Nitsolova 2008: 248).

II.2.1. The morphological categories person, number and gender of the verbs in Bulgarian

The morphological categories of the verbs in Bulgarian that are subject of analysis in this study are: person, number, gender (in some cases), tense, type of action and taxis. The main subject of the study are the categories that construct the hypercategory tense with focus on the transmission of grammatical information during translation. The categories person, number and gender will be briefly presented.

The morphological category person of verbs characterizes the object (phenomenon) "to which the dynamic procedural attribute is assigned, in terms of its relation to the act of speech" (Nitsolova 2008: 225). In Bulgarian linguistic studies the main conception is that the category of person is a three-member category and its meaning its based on the opposition communication – non-communicator (Gerdzhikov 1984: 222). The feature communicator unites first and second person, and the third person has the non-communicator feature.

The category of number of verbs expresses "the characteristic of objective quantitative features of the phenomena" (objects), "to which the dynamic attribute is assigned, it is a

quantitative characteristic of the verb doer" (Nitsolova 2008: 225). In the synthetic and analytical forms of personal verbs expressing the grammatical category of tense there are clear morphological indicators for distinguishing between singular and plural number.

Some of the analytical singular verb forms that contain a past perfective active participle (or past passive participle) express an additional grammatical meaning of gender.

II.2.2.   Characteristics of the temporal system in Bulgarian

The "birth" of the theoretical description of the modern Bulgarian temporal system can be attributed to the doctoral dissertation of L. Andreychin (Andreychin 1938) and his later "Basic Bulgarian Grammar" (Andreychin 1944). In his researches L. Andreychin clearly describes the resultativity of the perfect tenses and relativity, as well as the relations between them and the morphological category of tense, Andreychin divides the tenses into absolute and relative and tenses with a simple and complex chronological relation (Andreychin 1944: 208). In later researches the characteristics of the described temporal system are redefined and reconceptualised.

The authors of the different concepts in recent times are unanimous about the true meaning of the morphological category of tense and its differentiation from other meanings borne by the Bulgarian verb forms. The actual meaning of the morphological category of tense can only be constructed in relation to a particular orientation moment (the moment of a speech or another moment) that can be considered as the formal "temporal centre" to which relations can be in the privative opposition of a-synchronicity - synchronicity (Kutsarov 2007: 244). Within the frames of the a-synchronicity the opposition of posteriority and anteriority is additionally constructed (Kutsarov 2007: 244). The morphological category of tense in Bulgarian consists of the following grammemes (Kutsarov 2007: 244):

-        the grammeme **Present tense** expresses simultaneity of the dynamic feature with the act of speech or lack of information for a-synchronicity;

-        the grammeme **Past tense** expresses anteriority of the dynamic feature in relation to the act of speech;

-        the grammeme **Future tense** expresses posteriority of the dynamic feature in relation to the act of speech.

According to the system presented by Iv. Kutsarov (Kutsarov 2007: 244) in Bulgarian information for simultaneity (or lack of information about a-synchronicity) of the action and

the act of speech is expressed by the forms of Praesens, Imperfect (together with relativity), Perfect (together with resultativity) and Plusquamperfectum (together with relativity and resultativity). Information for posteriority of an action in relation to the act of speech is expressed by the forms of Futurum, Futurum Exactum (together with resultativity), Futurum Praeteriti (together with relativity) and Futurum Exactum Praeteriti (together with relativity and resultativity). Information for Anteriority of an action in relation to the act of speech is expressed by the forms of Aorist.

The semantics of the morphological category type of action is based on "the relation between action and result of action at the same orientation moment" (Kutsarov 2007: 249). There are different conceptions about the type of the categorical opposition: equipolar (resultativity – action) (Gerdzhikov 1984: 115) or privative (resultativity - non-resultativity) (Kutsarov 2007: 250). Information about the type of action in Bulgarian can express the following grammemes (Kutsarov 2007:250):

- the grammeme **resultativity** expresses a result of action, actualisation of action in the closest sequential orientation moment;

- the grammeme **non-resultativity** expresses lack of information for a result of action, and in its main meaning expresses action.

According to the system presented by Iv. Kutsarov (Kutsarov 2007: 250 – 253) in Bulgarian information for resultativity is expressed by the forms of Perfect, Futurum Exactum, Plusquamperfectum (together with relativity to a past orientation moment) and Futurum Exactum Praeteriti (together with information for posteriority of a result and relativity to a past orientation moment). Non-resultativity is expressed by the forms of Praesens, Aorist, Imperfect, Futurum and Futurum Praeteriti.

The third category which compiles the grammatical category (hypercategory) tense in Bulgarian is the category of **taxis** (relativity) and expresses "the relations between actions and results from actions that are oriented to an orientation moment in the past (Past Tense or other lexical equivalent) and actions that are not orientated to an orientation moment in the past" (Kutsarov 2007: 255). Iv. Kutsarov points out based on the opinion of G. Gerdzhikov that "the Aorist is the only past tense in the Bulgarian language" (Kutsarov 2007: 229) that the relative forms are oriented to the Aorist or another Past Tense signalisator, while the non-relative forms are not oriented or are oriented to the moment of speech.

While the morphological category of tense is constructed on the relation of an action to the moment of speech, the category of taxis is based on the relation between an action to another action (orientation moment). In other words, the category of taxis is based on the privative opposition of relativity – non-relativity (dependent - independent taxis) (Kutsarov 2007: 255). Information for relativity to a past orientation moment or lack of such information can be expressed by the following grammemes (Kutsarov 2007: 255):

- the grammeme **relativity** expresses actions (or results of actions) that are oriented to the single past tense in Bulgarian (Aorist) or other signalisator of the past orientation moment;

- the grammeme **non-relativity** expresses actions (or results of actions) that lack information for relation to a certain orientation moment or (in its main meaning) that are oriented to the moment of speech (absoluteness).

According to the system presented by Iv. Kutsarov (Kutsarov 2007: 253 – 257) in Bulgarian information for relativity is expressed by the forms of Imperfect, Futurum Praeteriti, Plusquamperfectum (together with resultativity) and Futurum Exactum Praeteriti (together with information for posterior resultativity in relation to a past utterance – past orientation moment). Lack of information for relativity is expressed by the forms of Praesens, Aorist, Futurum, Futurum Praeteriti and Perfect.

II.2.3.   Semantic characteristics of the Bulgarian temporal system.

The semantic model of H. Reichenbach (Reichenbach 1947) will be used to present the systematic relations and formal meanings of the members of the different morphological categories. Presenting the temporal relations in Bulgarian with by the model of H. Reichenbach is not a new approach. The system proposed by R. Nitsolova is used. The system introduced the following indications for the elements of the moment of utterance, the reference interval, the event interval and the moment in which the result of the event is existent (Nitsolova 2008: 264 - 265):

T_0 – moment of utterance (act of communication);

R – reference interval, whereas it is possible to have more than one reference interval (R_1);

E – event interval;

T_{res} – the moment in which the result of the event is existent.

The relations between these elements are marked as follows:

X < Y – X precedes Y, X is before Y;

X ∈ Y – X is an element from Y;

X ∉ Y – X is not an element from Y;

X ⊃ Y – X includes Y;

X ⊇ Y – X includes or coincides (is simultaneous) with Y.

Based on these elements and the relations between them, a formal-semantic description of the Bulgarian temporal system is given in Table 2.

Praesens	$T \in R; E \supseteq R$ For Praesens in Bulgarian the event interval includes reference interval which includes the moment of utterance.
Aorist	$R < T; R \supseteq E$ The semantic of Aorist expresses and event interval that is simultaneous with a reference interval in the past and it is enclosed in this interval.
Imperfect	$R < T; E \supset R$ Imperfect expresses an event interval that includes in it a reference interval in the past.
Futurum	$T < R; R \supseteq E$ The semantic of Futurum in Bulgarian expresses an action that it is taking place in a reference interval in the future situated after the moment of utterance.
Perfect	$T \in R; E < R, T_0; t_{res} \in R$ Perfect expresses a meaning that presents an event interval in the past, precedent to the reference interval that includes in it the moment of utterance.
Plusquamperfectum	$R < T; E < R < T_0; t_{res} \in R$ The semantic of Plusquamperfectum expresses an event interval that is taking place in the past before a certain past reference interval.
Futurum Exactum	$T < E < R; t_{res} \in R$ Futurum Exactum expresses an event interval that took place before a reference interval in the future in which the result of the action will be existent.
Futurum Praeteriti	$R < E < T$ Futurum Praeteriti expresses an event interval that is upcoming in relation to a reference interval in the past.
Futurum Exactum Praeteriti	$R < E < R_1 < T; t_{res} \in R_1$ For constructing the meaning of Futurum Exactum Praeteriti two reference interval are used. One of these intervals R is in the past and the event is upcoming in relation to it. Because Futurum Exactum Praeteriti is a resultative tense, the result of the action is existent in the second reference interval R1.

Table 2. *Formal semantical description of the Bulgarian temporal system.*

II.3. Characteristics of the English verbs

The English language belongs to the German language group and although some of its features that are common with other languages of the group are disturbed by later intense

linguistic influences, especially Norman, many common features have remained in the modern English language. Unlike the Bulgarian language, in which the theoretical description of the language categories begins relatively late, the English linguistics inherits the age-old tradition of the classical grammars built on the classical medieval languages – Latin and Greek. The development of modern linguistic studies from the end of the 19th and throughout the 20th century is marked by numerous scientific fields and schools that give a lasting impression on the formal and functional presentation of language, and English is usually the main subject of consideration when presenting different scientific concepts.

In English the verbs are divided into formal lexical-semantic classes and possess certain morphological categories (Rankova and Ivanova 1980: 88). The main lexical classes in which English verbs are divided based on their functional characteristics are auxiliary and main verbs. The verbs in English can be divided to transitive and intransitive verbs according to the semantics of the expressed action: whether the action directly affects an object that is syntactically realised as a direct object or not (Valeika and Buitkiene 2003: 89). An essential division of the verbs in English groups them into two main grammatical classes according to the formal means of forming past tense and past participle (Murphy 1991: 23). According to this division the verbs in English can be divided to regular and irregular verbs. Past tense and past participle forms of the regular verbs are formed by addition of inflection - (e)d to the infinitive. The irregular verbs form Past tense and Past participle according to different word-formation types, that can be divided to subtypes (Murphy 1991: 26).

The morphological categories that are subject to analysis in this research are person, number, tense and aspect. The grammatical category of gender is not a verb category in English and the meaning of "pedigree" is not expressed through the verb form.

II.3.1.   The morphological categories person and number of the verbs in English

In terms of the categories of person and number, the English language has lost much of its inflections that express the relation of the doer to the utterance. These categories exist in the English language because verb forms are necessarily combined with a subject expressed by a noun phrase or a pronoun. Only the verb form for 3rd person, singular, Present Simple Tense carries a formal marker for person and number of the synthetic forms (Filipova 2002: 27). Of the analytical forms only the auxiliary verbs in 1st and 3rd person, singular, Past Continuous

Tense and Past Continuous Tense; for 3rd person, singular, Present Perfect Continuous Tense and Past Perfect Tense carry formal marker for person and number (Filipova 2002: 67).

In this research it is assumed that the explicit subject or the personal pronouns are not part of the verb forms in English.

II.3.2. Characteristics of the temporal system in English

The first theoretical descriptions of the English temporal system are significantly influenced by the Latin grammars. At first, there were three tenses defined in English (Johnson 1640: 21), with later descriptions of the tense system adhering to this notion and naming the tenses by their Latin names. R. Lowth was the first to mention Perfect and Imperfect temporal forms (Lowth1762: 34). The conception of a three or two-membered tense category with two or three-membered the aspect category is acknowledged in the second half of the 19th and beginning of the 20th centuries. Views on the twelve-membered temporal system emerged in the early 20th century with clearly defined inventory, but with different names. For the first time in the grammar of Baskervill and Sewell 12 temporal forms - 3 tenses with 4 aspects were listed (Baskerville and Sewell 1895: 152). There are also more complicated modern systems that describe basic and secondary times, aspect forms and modal forms. For the purposes of this research, we acknowledge the notion of the grammatical category of tense in English to be a hypercategory constructed by the morphological categories of time and aspect.

Most contemporary models of the verb morphological categories in English adhere to the view of 3 basic tenses - past, present and future (House and Stevens 2000; Lewis and Mol 2013). There are also views about 4 basic tenses - past, present, future and future in the past or past future tense (Rankova and Ivanova 1980; Filipova 2002; Kabakchiev 2003). The differential features of the morphological category tense in English are two - anteriority and posteriority (Kabakchiev 2003: 132). The Present Tense does not contain information for posteriority or anteriority of the action. Information for anteriority contains the Past Tense and information for posteriority - the Future Tense (and Future Tense in the Past, that contains information for posteriority of the action in relation to a past orientation moment).

The aspect in English is a complex morphological category that relates the continuity or perfectivity of an action to an external (relative to the action itself) orientation moment (Valeika and Buitkiene 2003: 84). The action can be continuous, perfective, without information about continuity or perfectivity, or it can be perfect continuous in respect to each of the orientation moments of the four basic tenses. These relations are independent of the

lexical meaning of the verb (Filipova 2002: 42). Some authors decompose the morphological category of aspect in English into two separate categories – the aspect category (Valeika and Buitkiene 2003: 84), which expresses the continuity or lack of information about continuity of the action relative to a given orientation moment, and the category – order that relates the perfectivity of an action to a particular orientation moment — orders the moment of the action before the orientation moment (Valeika and Buitkiene 2003: 97). Despite that in the linguistics studies it is accepted that the category of aspect is a single complex category, constructed by the equipolar opposition of continuity – perfectivity (Rankova and Ivanova 1980; Filipova 2002; Kabakchiev 2003). The differential characteristics of the category of aspect in English are 2: continuity and perfectivity (perfectivity / resultativity) (Kabakchiev 2003: 134). No information about continuity of perfectivity is expressed by the Simple tenses (Filipova 2002: 69). The Continuous tenses express the meaning of continuity (Filipova 2002: 82) and the Perfect tenses - the meaning of perfectivity (Filipova 2002: 90). Both of the meanings are being expressed by the Perfect Continuous Tenses (Filipova 2002: 101).

II.3.3. Semantic characteristics of the English temporal system.

A description of the English temporal system is presented with the system of H. Reichenbach. For the formal description the elements and relations accepted for the Bulgarian language system are used in order for a formal-semantic comparison between Bulgarian and English to be possible.

A formal-semantic description of the English temporal system is given in Table 3.

Present Simple Tense	$E=R=T$ In the semantics of Present Simple Tense the moment of utterance, the moment of reference, and the moment of action coincide completely, with no information about the completion or repetition of the action.
Present Continuous Tense	$T = R; E \supseteq R$ The semantics of Present Continuous Tense expresses an action, which moment includes the moment of reference, which coincides with the moment of utterance.
Present Perfect Tense	$E < T = R; t_{res} \in R$ Present Perfect Tense expresses a past action that is before the moment of utterance. Present Perfect Tense is a resultative tense and the result from the action is existent in the moment of utterance, which coincides with the moment of reference.
Present Perfect Continuous Tense	$T \in R; E < R; t_{res} \in R$ The semantics of Present Perfect Continuous Tense is based on a moment of action, which is before the moment of reference and the moment of utterance, which coincide. The result from the action is existent within the moment of reference with no information for completion of the action.

Past Simple Tense	$E = R < T$ The semantics of Past Simple Tense expresses the moment of the action, which coincides with a past moment of reference, which is before the moment of utterance.
Past Continuous Tense	$R < T; E \supset R$ For Past Continuous Tense the moment of reference is before the moment of utterance, and the moment of the action passes through it – the beginning of the action is before the moment of reference and there is no information for completion of the action in relation to the moment of utterance.
Past Perfect Tense	$E < R < T; t_{res} \in R$ The semantics of Past Perfect Tense in English is based on a moment of the action, which is anterior to a past moment of reference and the result from the action is existent in the moment of reference.
Past Perfect Continuous Tense	$R \supseteq E; R < T; t_{res} \in R$ Past Perfect Continuous Tense expresses a moment of the action that is anterior to the moment of reference and the result from the action is existent in the moment of reference. The verb forms for this tense do not express information for termination of the event interval within the reference interval or before it.
Future Simple Tense	$T < R = E$ The semantics of Future Simple Tense is based on a moment of reference in which is the moment of the action. The moment of reference is after the moment of utterance.
Future Continuous Tense	$T < R; R \supseteq E$ For Future Continuous Tense the moment of reference is posterior to the moment of utterance and the moment of the action is included in the moment the moment of reference.
Future Perfect Tense	$T < E < R; t_{res} \in R$ For Future Perfect Tense the moment of the action is posterior to the moment of utterance and the result from the action is existent in a future reference interval that is after the moment of utterance and the moment of the action.
Future Perfect Continuous Tense	$R \supseteq E ; T < E; t_{res} \in R$ The semantics of the Future Perfect Continuous Tense is based on a moment of reference that is posterior to the moment of utterance and the result from the action is existent in the moment of reference. The moment of the action is posterior in regards to the moment of utterance.
Future Simple Tense in the Past	$R < E < T$ For Future Simple Tense in the Past the moment of the action is posterior to a moment of reference in the past.
Future Continuous Tense in the Past	$T > R > R_1; E \supset R_1$ For Future Continuous Tense in the Past in English both of the moments of reference are before the moment of utterance. The first moment of reference is anterior to the second one and the moment of the action is posterior to the first moment of reference and terminates before the second moment of reference - the beginning of the action is before the second moment of reference and there is no information for termination of the action within the second moment of reference.
Future Perfect Tense in the Past	$R < E < R_1 < T; t_{res} \in R_1$ The semantics of Future Perfect Tense in the Past is based on two subsequent moments of reference anterior to the moment of utterance. The moment of the action is between the two moments of reference and it is posterior to the first one. The result from the action is existent within the second moment of reference.

Future Perfect Continuous Tense in the Past	$R < E < R_1 < T$; $E \supset R_1$ $t_{res} \in R_1$ The semantics of Future Perfect Continuous Tense in the Past is based on two subsequent moments of reference anterior to the moment of utterance. The moment of the action is posterior to the firs moment of reference and reaches and/or encompasses the second moment of reference and the result from the action is existent in the second moment of reference or there is no information within the second moment of reference for termination of the action.

Table 3. *Formal semantical description of the English temporal system.*

II.4. Comparative analysis of the characteristics of the verb systems and the word formation of temporal forms in Bulgarian and English

The formal correspondences between the temporal categories in Bulgarian and English are derived on the basis of the basic use of specific tenses, which is determined by the positions of the moment of utterance (T), the moment at which the result of action is existent (T_{res}), the intervals / moments of reference (R, R_1, R_2) and the interval / moment of action (E) on the temporal axis. The comparison aims to present the basic usage of the tenses. The formal analysis will also be limited to presenting the relations between the elements T, T_{res}, R, and E. A comprehensive description of all the possible transpositional meanings and formal concurrences between their meanings is beyond the scope of this research. The relations between the moment of utterance, the moment of action/event interval and the moment of reference/reference interval can be presented as follows:

– a relation of coinciding between the moment of utterance, the event interval and the reference interval (E=R=T) in both languages is expressed by the semantics of Praesens in Bulgarian and Present Simple Tense and Present Continuous Tense in English. The main difference is in the characteristics of the event interval in relation to the reference interval. In Bulgarian in the semantics of Praesens the event interval is included in the reference interval, while in the semantics of Present Simple Tense in English there is a total coincidence between the event interval and the reference interval;

– a relation of coincidence between the event interval with the reference interval and relation of anteriority of the reference interval in relation to the moment of utterance (E=R<T) is expressed by Aorist and Imperfect in Bulgarian and Past Simple Tense and Past Continuous Tense in English. The semantics of Imperfect in Bulgarian expresses an event interval that includes a reference interval in the past and Aorist – the event interval and the reference interval in the past coincide. In the semantics of Past Simple

Tense in English the event interval coincides with a reference interval in the past, whereas in the semantics of Past Continuous Tense the moment of reference is anterior to the moment of utterance and the moment of the action passes through it;

− a relation of coincidence between the reference interval and the moment of utterance and relation of posteriority of the reference interval in relation to the moment of utterance ($E=R>T$) is expressed by Futurum in Bulgarian and Future Simple Tense and Future Continuous Tense in English. In the semantics of Future Continuous Tense the event interval passes through the reference interval and there is no information for termination of the action within the reference interval.

The relations between the moment of utterance, the moment of the action/event interval, the moment of reference/reference interval and the moment in which the result of the action is existent can be presented as follows:

− a relation of coincidence between the reference interval, moment of utterance and the moment in which the result of the action is existent and a relation of anteriority of the event interval ($E<R=T=T_{res}$) is expressed by Perfect in Bulgarian and Present Perfect Tense and Present Perfect Continuous Tense in English. In the semantics of Present Perfect Continuous Tense there is information for anteriority of the beginning of the event interval in relation to the reference interval, but there is no information for termination.

− a relation of anteriority of the event interval in relation to the reference interval and relation of anteriority of the reference interval in relation to the moment of utterance and coincidence in relation to the moment in which the result of the action is existent ($E<R=T_{res}<T$) is expressed by Plusquamperfectum in Bulgarian and Past Perfect Tense and Past Perfect Continuous Tense in English. In the semantics of Past Perfect Continuous Tense there is information for anteriority of the beginning of the event interval in relation to the reference interval, but there is no information for termination.

− a relation of anteriority of the event interval in relation to the reference interval and relation of posteriority of the reference interval in relation to the moment of utterance and coincidence in relation to the moment in which the result of the action is existent ($E<R=T_{res}>T$) is expressed by Futurum Exactum in Bulgarian and Future Perfect Tense and Future Perfect Continuous Tense in English. In the semantics of Future Perfect Continuous Tense there is information for anteriority of the beginning

of the event interval in relation to the reference interval, but there is no information for termination.

A relation between the presented elements above – the moment of utterance, the moment of action/event interval, the moment of reference/ reference interval, the moment in which the result from the action is existent and a second reference interval – in both languages can be presented as follows:

−   a relation of posteriority of the event interval in relation to a first reference interval, which is anterior in relation to a second reference interval, which in turn is posterior to the moment of utterance $(R>E>R_1>T)$ is expressed by Futurum Praeteriti in Bulgarian and Future Simple Tense in the Past and Future Continuous Tense in the Past in English. In the semantics of Future Continuous Tense in the Past there is information for anteriority of the beginning of the event interval in relation to the second reference interval, but there is no information for termination;

−   a relation of posteriority of the event interval in relation to a first reference interval, which is anterior in relation to a second reference interval, which in turn is posterior to the moment of utterance and in relation of coincidence to the moment in which the result from the action is existent $(R>E>R1=T_{res}>T)$ is expressed by Futurum Exactum Praeteriti in Bulgarian and Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past in English. In the semantics of Future Perfect Continuous Tense in the Past there is information for anteriority of the beginning of the event interval in relation to the second reference interval, but there is no information for termination.

In regards of the grammatical meanings of the categories that construct the temporal meanings in Bulgarian and English the following conclusions can be drawn:

-        verb forms that express simultaneity or no information for a-synchronicity of the action in relation to the moment of utterance without other meanings are existent in both languages. In Bulgarian these are the verb forms for Praesens, and for English – the verb forms for Present Simple Tense and Present Continuous Tense (together with information for continuity of the action);

-        verb forms that express information for posteriority of the action in relation to the moment of utterance without other meanings are existent in both languages. In Bulgarian these are the verb forms for Futurum, and for English – the verb forms for

Future Simple Tense and Future Continuous Tense (together with information for continuity of the action);

-        verb forms that express information for posteriority of the result from the action in relation to the moment of utterance are existent in both languages. In Bulgarian these are the verb forms for Futurum Exactum, and for English – the verb forms for Future Perfect Tense and Future Perfect Continuous Tense (together with information for continuity of the action);

-        verb forms that express anteriority of the action in relation to the moment of utterance are existent in both languages. In Bulgarian these are the verb forms for Aorist, and for English – the verb forms for Past Simple Tense.

-        verb forms that express simultaneity of the result from the action in relation to the moment of utterance are existent in both languages. In Bulgarian these are the verb forms for Perfect, and for English – the verb forms for Present Perfect Tense and Present Perfect Continuous Tense.

-        verb forms that express existence of the result from the action in relation to a moment in the past are existent in both languages. In Bulgarian these are the verb forms for Plusquamperfectum, and for English – the verb forms for Past Perfect Tense and Past Perfect Continuous Tense.

-        verb forms that express posteriority of the action in relation to an orientation moment in the past are existent in both languages. In Bulgarian these are the verb forms for Futurum Praeteriti, and for English – the verb forms for Future Simple Tense in the Past and Future Continuous Tense in the Past;

-        verb forms that express posteriority of the result from the action in relation to an orientation moment in the past are existent in both languages. In Bulgarian these are the verb forms for Futurum Exactum Praeteriti, and for English – the verb forms for Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past;

The formal equivalents between the Bulgarian and English verb forms, as well as the number of possible forms is presented in Table 4.

Bulgarian	English	Number of Forms
Praesens	**Present Simple Tense/Present Continuous Tense**	
1 person, singular; 2 person, singular; 3 person, singular; 1 person, plural; 2 person, plural; 3 person, plural;	3 person, singular; 1/2 person, singular/1/2/3 person, plural/ 1 and 3 person, singular; 2 person, singular/1/2/3 person, plural	6:2/6:3
Aorist	**Past Simple Tense**	
1 person, singular; 2/3 person, singular; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	5:1
Imperfect	**Past Continuous Tense**	
1 person, singular; 2/3 person, singular; 1 person, plural; 2 person, plural; 3 person, plural;	1 and 3 person, singular; 2 person, singular/1/2/3 person, plural	5:2
Perfect	**Present Perfect Tense/Present Perfect Continuous Tense**	
1 person, singular, masculine; 1 person, singular, feminine; 1 person, singular, neutral; 2 person, singular, masculine; 2 person, singular, feminine; 2 person, singular, neutral; 3 person, singular, masculine; 3 person, singular, feminine; 3 person, singular, neutral; 1 person, plural; 2 person, plural; 3 person, plural;	3 person, singular; 1/2 person, singular/1/2/3 person, plural	12:2
Plusquamperfectum	**Past Perfect Tense/Past Perfect Continuous Tense**	
1 person, singular, masculine; 1 person, singular, feminine; 1 person, singular, neutral; 2/3 person, singular, masculine; 2/3 person, singular, feminine; 2/3 person, singular, neutral; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	9:1
Futurum	**Future Simple Tense/Future Continuous Tense**	
1 person, singular; 2 person, singular; 3 person, singular; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	6:1
Futurum Exactum	**Future Perfect Tense/Future Perfect Continuous Tense**	
1 person, singular, masculine; 1 person, singular, feminine; 1 person, singular, neutral; 2 person, singular, masculine; 2 person, singular, feminine; 2 person, singular, neutral; 3 person, singular, masculine; 3 person, singular, feminine; 3 person, singular, neutral; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	12:1
Futurum Praeteriti	**Future Simple Tense in the Past**	
1 person, singular; 2 person, singular; 3 person, singular; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	6:1
Futurum Exactum Praeteriti	**Future Perfect Tense in the Past/Future Perfect Continuous Tense in the Past**	
1 person, singular, masculine; 1 person, singular, feminine; 1 person, singular, neutral; 2 person, singular, masculine; 2 person, singular, feminine; 2 person, singular, neutral; 3 person, singular, masculine; 3 person, singular, feminine; 3 person, singular, neutral; 1 person, plural; 2 person, plural; 3 person, plural;	1 form for person and number	12:1

Table 4. *Relation between the forms containing information for the grammatical category of tense in Bulgarian and English.*

III.   Creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms"

This chapter presents a brief overview of the evolution of machine translation as a scientific field and the current theories and practices in the creation of linguistic resources for the purposes of statistical language modelling and statistical machine translation. The main result of the research is presented – "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

III.1.   Machine translation – goals and historical development

The term machine translation can be presented as the practical application of computers to translate texts from one natural language to another (Hutchins 1995: 431). The scientific approaches and methods used in machine translation can be divided into three major types that follow the historical development of the scientific field. The first (in historical context) type is the so-called "direct translation" approach. In the direct translation approach the machine translation systems are specifically designed for a specific pair of working languages (Hutchins 1995: 433). The systems for direct translation are usually constructed of handwritten rules for analysing, generating and converting texts and large bilingual dictionaries. The second type of machine translation systems – the transfer-based machine translation systems emerged as a successor of the direct approach translation systems. The transfer-based systems consistently upgrade the various linguistic levels of the source language to construct an intermediate abstract representation between the source and target languages. With accumulation of computing power and development of modern perceptions of the nature of language, the systems for statistical language modelling emerged. Statistical language modelling takes into account the final products of the language in order to examine its natural processes and behaviour (Cohen 2010: 74). The aim of statistical machine translation is to represent the likelihood of two language units in two or more different languages to correspond to each other. Statistical machine translation systems can be described as a function that maximizes the product of logical and semantic correspondence of a sentence (S and T) in the source and target language and the fluency of T with respect to the target language. In the 21st century the neural machine translation is emerging. The basic structure of neural machine translation consists of two repetitive layers of neurons. The input layer (the analyser) analyses the variable length input text units and encodes them into a fixed length vector. The second neural layer (the

decoder) generates target text units based on the vector generated by the first network (Firat et al. 2016: 866).

Both traditional statistical translation systems and neural machine translation systems rely primarily on sufficiently large and reliable language corpora to be trained and evaluated.

III.2.   Language corpora as a resource for machine translation

According to the definition given by the European Linguistic Resources Association the term linguistic resource means a system of linguistic data or descriptions of such data that is available for machine processing for various purposes. Language resources may be classified according to the structure in which the language data is presented or according to their specific content. Depending on how they are structured, language resources can be divided into thesauri, corpora, dictionaries, thematic dictionaries, terminology bases or databases. Corpora are a collection of electronic texts, selected by external criteria, to represent, as possible, a language or a variety of languages and to provide data for linguistic research (Sinclair 2005 – ref. by. Koeva 2014: 29). Definitions for language corpus usually emphasize the criteria for selection of corpora units, the importance of the corpora for various linguistic studies, and the opportunities that corpora provide for computer processing (Koeva 2014: 29).

Enclosures are presented through the traditionally accepted features of the language corpus – representativeness, balance and sampling (Leech 1991: 10). The representativeness of a language corpus is determined by the extent to which it can cover the full range of different manifestations of a given language phenomenon. The balance of a corpus depends on its text typology – the classification and the characteristics of the text types involved. The sampling of a language corpus is related to its representativeness and balance.

III.3.   Motivation for the corpus-based approach in the research of the Bulgarian-English grammatical parallels

The temporal systems in Bulgarian and English share a basic feature – they consist of categories that are structurally organized and can be considered as members of a hypercategory. The greater number of possible morphological categories, and respectively, the greater number of possible grammatical meanings in Bulgarian contributes to the greater ambiguity during translation due to the fact that the grammaticalised information from the source language must be reduced or unevenly distributed between different grammatical target language categories.

However, the characteristics of the Bulgarian and English temporal forms have very comparable characteristics (Lazarov 2016a): both languages have the morphological category of tense, which expresses the relation of the action to the moment of utterance; in Bulgarian and English there is a morphological category that expresses the relation of the result of an action to an orientation moment – in Bulgarian this is the category of the type of action (resultativity), and in English this is the complex morphological category of aspect, which can express continuity and / or perfectivity; the word formation of the two languages has auxiliary verbs which carry some of the grammatical meaning of the verb form; morphological categories that construct temporal meanings in both languages can be considered as part of a hypercategory.

The greater number of possible grammatical categories, and respectively, the greater number of possible grammatical meanings in Bulgarian contributes to the greater ambiguity during translation due to the fact that the grammaticalised information from the source language must be reduced or unevenly distributed between different morphological target language categories (or to be realised on lexical or syntactic level). Due to the characteristics of the two languages, a significant amount of the grammaticalised information is transmitted through different language means and on different linguistic levels in the semantic transfer during translation from Bulgarian to English.

One of the goals of the research is the creation and description of a translation model containing information about the translation equivalents of the Bulgarian verb forms in English. In order to accomplish this goal, it is necessary to: create a sufficiently large parallel corpus annotated with the morphological information for the verb forms in Bulgarian and their translation equivalents in English; to extract and analyse the statistics from the annotated corpus and to perform experiments with translation models that are enriched with the annotated data from the corpus in order to verify its practical application. The analysis of the frequency of the verb forms in Bulgarian and their translation equivalents in English provides an opportunity to research the features of the temporal categories and the transfer of the information they express during translation.

III.4. "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

The "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is created in the period 2017 – 2018 and his goals is to be a resource for exploring the grammatical parallels

between the verb systems in Bulgarian and English: the realisation of the morphological categories that construct the temporal systems in both languages.

The goals of the creation of the Corpus are in two directions – theoretical and practical. The theoretical goals are to extract and correlate objective quantitative and distributive language data for the temporal systems in Bulgarian and English and comprehensive (for the volume of the Corpus) qualitative and quantitative language manifestations for reliable analysis. The practical goals are to provide credible linguistic data to support the creation of various language applications for computer processing of the natural language, such as a statistical annotation model and a statistical translation model for verb form.

III.4.1.   Selection of language resources included in the Corpus

The "Bulgarian-English Parallel Corpus with Aligned Verb Forms" contains samples from other language resources that are selected according to pre-developed criteria.

The main criterion for selecting language resources to be included in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is the presence of qualitative morphological annotation, since it is an essential part of the processing of linguistic resources for statistical data extraction and construction of language models. The evaluated resources are the Bulgarian National Corpus, the "Bulgarian-English Parallel Corpus" (Koeva et al. 2012a) and the "Bulgarian-English Sentence- and Clause-Aligned Corpus" [1] (Koeva et al. 2012b). The "Bulgarian-English Sentence- and Clause-Aligned Corpus" has the following characteristics:

-       a detailed morphological monolingual and multilingual annotation by the Bulgarian Language Processing Chain [2];

-       a multilingual annotation that includes sentence alignment in both languages (and clauses) and the alignment is manually corrected, and the auto-alignment verification and correction are performed with a specially designed program.

The criteria for the selected language resources can be divided as follows:

-       according to the characteristics for a language corpus: the resources must be representative of the linguistic phenomenon (the specifics of the verb usage predefines the unbalance of the text types to be included in the Corpus);

---

[1] https://dcl.bas.bg/bulenac/
[2] https://dcl.bas.bg/webservices/

- according to the current state of resources: another basic requirement is that the resource should represent current use of the language;

- according to the linguistic information in the available resources: the appropriate resources should represent parallel Bulgarian-English aligned texts at a certain level – sentence or paragraph aligned, the appropriate resources should also have an annotation layer with morphological information – markers for part of speech and grammatical characteristics;

- according to the extralinguistic characteristics of the resources: the resource meta information must indicate which language is the original and the language of the translated text (direction of translation in the resource) [3];

- according to copyright: the selection also determines the copyright status of the resources: they must be copyright-exempt, copyrighted for work, or always cited.

According to the presented criteria the choice of resources is limited to the "Bulgarian-English Sentence- and Clause-Aligned Corpus". In order to ensure the diversity of the included temporal forms, copyrighted bilingual documents are collected from the Internet and also examples from "Bulgarian Grammar" (Nitsolova 2017) are used.

III.4.2. Structure of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms"

The Corpus includes selected parts of the "Bulgarian-English Sentence- and Clause-Aligned Corpus", Internet documents, and examples from "Bulgarian Grammar", that are paragraphs or sentences. The meta information of each corpus unit includes data for the source (file name or URL), source and target language, date of incorporation into the Corpus (November 2018 – January 2019), existence of part of speech and grammatical characteristics tags, and information about the language domain (fiction texts, journalistic texts, scientific texts, subtitles and examples from "Bulgarian Grammar").

In its present state the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" consists of 84% of texts from the "Bulgarian-English Sentence- and Clause-Aligned Corpus", 15.5% of texts from the Internet and 0.5% of sentences extracted from examples from

---

[3] Due to insufficient amount of this information, this criterion is not systematically applied during the creation of the Corpus.

Bulgarian Grammar. In order to mention the copyright of the examples from "Bulgarian Grammar", source information is included in the metadata of the respective files.

During the work on the design of the Corpus, one of the main methods for achieving balance of a language resource was adopted – basing it on an already adopted structure and extracting random samples from a larger corpus. According to external criteria, a limited number of texts were selected from the Internet, as well as specially designed sentences for the purpose of completeness of the data in the Corpus. After each new addition of language material in the Corpus, the ratio between the total number of Bulgarian verb forms and the illustrated temporal meanings is extracted in order to avoid the increase of examples involving the same temporal forms. For this reason, the taxonomy of text types in the Corpus does not have a unified structure. The textual taxonomy is based on a periodic checking on the linguistic saturation of the Corpus – whether the addition of a new processed text to it, will change significantly the number of linguistic units (verb forms) or it will remain relatively constant.

The low frequency of some temporal forms in Bulgarian implies that during the phases of adding language material to the Corpus and reaching linguistic saturation, the estimated assumed values for the frequency of these temporal forms will not change significantly and will remain low throughout the whole process. In other words, when a new linguistic data specially selected by certain criteria is incorporated and introduces new occurrences of temporal forms to the array, it cannot be expected that the number of new occurrences will continue to change proportionately in the next stages of operation. In this sense the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is an unbalanced linguistic corpus.

III.4.3. Annotation conventions during the creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms"

The linguistic data in the Corpus consists of two layers of annotation. The first annotation layer is the layer of part of speech tags, which consists of markers that show part of speech values for a given word. The TreeTagger software was used to annotate texts from the Internet and examples from "Bulgarian Grammar", and also to unify the tags with the "Bulgarian-English Sentence- and Clause-Aligned Corpus". The second annotation layer contains information about the analytical verb forms and the correlation of temporal categories in the two languages. The freely available WebAnno language tool was used for the addition of the second layer of annotation.

The inventory of the second annotation layer consists of 26 values that represent the grammatical category of tense. The following principles are adopted when implementing the tags of the second annotation layer:

- the linguistic markers for Bulgarian and English represent the classical members of the hypercategory of tense in both languages – 9 for Bulgarian and 17 for English (16 for the temporal forms and 1 for the infinitives);

- the linguistic markers of the second annotation layer do not contain information about the grammatical person, number or gender of verb forms, since this information is formed by the values of the markers for the morphological characteristics of the words;

- it is permissible for the aligned forms not to be equally distributed in both languages – it is possible that a verb form in Bulgarian may not have the corresponding equivalent in English (the translation is with other parts of the speech) and vice versa;

- the markers comprise only of the temporal characteristics of the forms, they do not contain information for other grammatical categories such as conditional, voice, evidentiality, positive or negative forms. Negative forms or passive voice forms are also marked, but only their temporal characteristics are considered.

The second annotation layer was added manually and / or semi-automatically with the WebAnno application, which allows full automatically alignment and tagging with linguistic information based on a probabilistic model and subsequent correction.

The Corpus is in CoNLL[4] format. The sentences consist of one or more rows that contain a word and the rows contain the following columns: sequence of the token; the token; a marker for the morphological characteristics of the word from the first annotation layer; a marker for the analytical verb form, if the token is part of such; verb form identification number. The identification number can be in three different relations between the two languages: the verb form identification number in Bulgarian corresponds to the verb form identification number in English; the identification number of the verb form in Bulgarian does not have the corresponding equivalent in English due to translation in English with other lexical means that are not a verb form; the verb form identification number in English does not have

---

[4] https://universaldependencies.org/format.html

the corresponding equivalent in Bulgarian due to the translation of non-verb parts of the Bulgarian sentence with verb forms in English.

III.5. Characteristics of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

In its current state the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is a small, unbalanced, sampled (for its intended purpose), bilingual parallel corpus with a total volume of 5,070 annotated and aligned verb forms and their translation correspondences. The Bulgarian texts are 45,843 words (48.12% of the total volume) and the English texts are 49,432 words (51.88% of the total volume) (Lazarov 2019b).

The schematic representation of the texts in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is as follows (Lazarov 2019b: 111):

- By language domain:

    - 65.5% fictional texts – stories, novels, tales, etc., of which 60% are from the "Bulgarian-English Sentence- and Clause-Aligned Corpus" and 40% are texts from the Internet.

    - 18% journalistic texts – news, announcements, press releases, etc., of which 100% are from the "Bulgarian-English Sentence- and Clause-Aligned Corpus";

    - 3% scientific texts – researches, articles, of which 100% are from the "Bulgarian-English Sentence- and Clause-Aligned Corpus";

    - 13% subtitles, of which 100% are from the "Bulgarian-English Sentence- and Clause-Aligned Corpus";

    - 0.5% other texts that are examples from "Bulgarian Grammar".

- By source and target language:

    - 40.1% of the text are original in Bulgarian with English translation;

    - 35.4% of the text are original in English with Bulgarian translation;

    - 24.5% of the text are original in a third language with Bulgarian and English translation.

- By source of the texts:

- 84% of the texts are from the Bulgarian-English Sentence- and Clause-Aligned Corpus;

- 15.5% of the texts are from internet resources;

- 0.5% of the texts are constructed sentences from examples from "Bulgarian Grammar".

The Corpus was created as part of project "Bulgarian-English grammatical parallels with respect to machine translation. Enriching statistical translation mode from Bulgarian to English with linguistic information" as part of the Support program for young scientists and PhD students at the Bulgarian Academy of Sciences.

IV. Analysis of the grammatical parallels in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

The main information that can be extracted from the Corpus is information about the usage of the temporal forms in the two languages separately and in comparative plan. This chapter of the research presents statistically-based analysis of the data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" about the distribution of the temporal forms in Bulgarian and English. The English translation equivalent of the Bulgarian temporal forms attested in the Corpus are also analysed.

IV.1. The "Bulgarian-English Parallel Corpus with Aligned Verb Forms" in the context of the research

Before presenting the statistics extracted from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" its main characteristics are discussed. A smaller experimental corpus was created before the creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" in order to conduct experiments on its design and working process (Lazarov 2017). This experimental study showed that the verb forms in Bulgarian and English have an uneven distribution in different types of texts and cannot be expected to demonstrate all possible manifestations in different contexts. On the other hand, it was confirmed that the corpus-based approach is suitable for extracting statistical information about the translation realisations of verb forms from Bulgarian to English, and that the presence of certain grammatical categories in Bulgarian as a source language predetermines the presence of certain grammatical categories in English as the target language.

The main information that can be extracted from the Corpus is information about the usage of the temporal forms in the two languages separately and in comparative plan. In the first annotation layer only the synthetic verb forms are annotated, including forms expressing tense (Praesens, Aorist, Imperfect for Bulgarian and Present Simple Tense and Past Simple Tense for English), conditional (indicative for both languages) and voice (active voice for both languages). The grammatical characteristics attributed in the second annotation layer refer to the analytical temporal forms in Bulgarian and English and do not contain information about other morphological categories. A main characteristic of the second annotation layer is that there is no correlation between verb forms in Bulgarian and other parts of speech that in English as translation equivalents (and vice versa).

IV.2. Statistically-based analysis of the data for the Bulgarian temporal forms in the Corpus

The frequency data for the verb forms in the Corpus is as follows: 2,437 verb forms in Praesens; 1,311 verb forms in Aorist; 710 verb forms in Imperfect; 269 verb forms in Perfect; 189 verb forms in Futurum; 164 verb forms in Plusquamperfectum; 47 verb forms in Futurum Praeteriti; 10 verb forms in Futurum Exactum and 5 verb forms in Futurum Exactum Praeteriti.

The following distribution is observed in regard to the temporal morphological categories in Bulgarian: 3,580 verb forms expressing a-synchronicity; 1,311 verb forms expressing anteriority; 251 verb forms expressing posteriority; 4,694 non-resultative verb forms; 448 resultative verb forms; 4,216 non-relative verb forms; 926 relative verb forms.

The statistical data shows that verb forms for Praesens are almost half of the annotated temporal forms for Bulgarian in the Corpus. The least frequent forms of Futurum Exactum Praeteriti are 0.1% of the annotated temporal forms for Bulgarian in the Corpus. It can be concluded that the non-marked members of the respective temporal morphological categories have higher distribution: a-synchronicity, non-resultativity and non-relativity. This peculiarity can be explained with the fact that the non-marked members can express their main meaning, but also the opposite meaning of the marked member and also in their general meaning they can express lack of information for the main meaning of the marked member of the category. Also it can be seen that 86.7% of the temporal verb forms in Bulgarian are synthetic verb forms and 13.3% are analytical verb forms.

IV.3. Statistically-based analysis of the data for the English temporal forms in the Corpus

The frequency data for the verb forms in the Corpus is as follows: 2,008 verb forms in Past Simple Tense; 1,430 verb forms in Present Simple Tense; 294 infinitives; 264 verb forms in Future Simple Tense; 159 verb forms in Past Perfect Tense; 117 verb forms in Past Continuous Tense; 103 verb forms in Future Simple Tense in the Past; 96 verb forms in Present Perfect Tense; 73 verb forms in Past Continuous Tense; 15 verb forms in Past Perfect Continuous Tense; 12 verb forms in Future Perfect Tense in the Past; 11 verb forms in Future Perfect Tense; 7 verb forms in Present Perfect Continuous Tense; 3 verb forms in Future Continuous Tense in the Past; 2 verb forms in Future Continuous Tense; there are now verb forms in Future Perfect Continuous Tense and Future Perfect Continuous Tense in the Past.

The following distribution is observed in regard to the temporal morphological categories in English: 3,805 verb forms of Simple tenses; 278 verb forms of Perfect tenses; 195 verb forms of Continuous tenses; 22 verb forms of Perfect Continuous tenses; 294 infinitives.

The statistical data shows that verb forms in Past Simple Tense are most frequent in the Corpus and there are no verb forms in Future Perfect Continuous Tense and Future Perfect Continuous Tense in the Past. The most frequent type of tenses in the Corpus are the Simple tenses (non-perfect, non-continuous) and the most complex tenses, such as the Perfect Continuous tenses have the lowest frequency in the Corpus. In English the second most frequent forms are the infinitives. From a semantic perspective the infinitive forms express the action on its own, without any relation to the verb form (Rankova and Ivanova 1980: 154), which implies that the necessary information for a complete translation cannot be transmitted on morphological level.

IV.4. Statistically based analysis of the translation of temporal forms from Bulgarian to English

The translation realisations in English of the Bulgarian temporal forms in the Corpus are presented in Table 5.

Tense in English / Tense in Bulgarian	Praesens	Aorist	Imperfect	Perfect	Futurum	Plusquamperfectum	Futurum Praeteriti	Futurum Exactum	Futurum Exactum Praeteriti
Present Simple Tense	1346	35	14	14	17	3	2	–	–
Present Continuous Tense	59	3	4	1	6	–	–	–	–
Present Perfect Tense	29	21	3	42	–	1	–	–	–
Present Perfect Continuous Tense	4	1	–	2	–	–	–	–	–
Past Simple Tense	289	1084	444	134	6	44	3	–	–
Past Continuous Tense	21	15	64	8	5	4	1	–	–
Past Perfect Tense	9	21	10	29	–	90	–	–	–
Past Perfect Continuous Tense	3	1	7	1	–	3	–	–	–
Future Simple Tense	161	1	1	–	102	–	–	–	–
Future Continuous Tense	–	–	–	–	2	–	–	–	–
Future Perfect Tense	–	–	–	–	–	–	–	10	–
Future Simple Tense in the Past	23	6	8	11	23	3	29	–	–
Future Continuous Tense in the Past	1	–	–	–	–	–	4	–	–

Future Perfect Tense in the Past	–	–	–	1	–	3	3	–	5
Other parts of speech	274	90	120	17	15	13	3	–	–
Infinitive	210	30	34	8	12	–	–	–	–

Table 5. *Translation realisations of the Bulgarian verb forms in English.*

The data presented in Table 5 shows that the verb forms for Praesens in Bulgarian have the most various translation realisations in English - 11 temporal forms, infinitives and other parts of speech. The significantly higher number of possible translation realisations of the verb forms in Praesens in English can be explained with the fact that in Bulgarian Praesens has the greatest number of possible transpositional usages (Nitsolova 2008: 274 - 276), some of which are impossible to translate with verb forms of Present tenses in English.

Despite the variety of translation realisations, the verb forms in Present Simple Tense in English have predominant frequency (Praesens in Bulgarian and Present Simple Tense in English have been considered as semantically equivalent in both languages based on formal-semantic analysis). Most limited in regard to the number of translation realisations are the verb forms in Futurum Exactum and Futurum Exactum Praeteriti – with only one corresponding temporal form. Within the data in the Corpus there are no translation realisations of Praesens in Bulgarian with verb forms in Future Continuous Tense, Future Perfect Tense, Future Perfect Continuous Tense, Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past in English.

The verb forms in Aorist in the Bulgarian texts of the Corpus are 1,311 or 25.50% of all the Bulgarian verb forms. The verb forms in Aorist are around 1/4 of all the verb forms in the Corpus and around 1/2 of the number of verb forms in Praesens.

The translation realisations of the verb forms in Aorist to English are diverse. Verb forms in Past Simple Tense in English have the highest frequencies, followed by translation realisations with other parts of speech. The verb forms in Aorist have a similar variety of translation realisations in English as the verb forms in Praesens. An explanation for this diversity can be sought both in the transpositional usages and in the principle of tense coordination in English.

Within the data in the Corpus there are no translation realisations in English of Aorist in Bulgarian with verb forms in Future Continuous Tense, Future Perfect Tense, Future Perfect Continuous Tense, Future Continuous Tense in the Past, Future Perfect Tense in the Past, Future Perfect Continuous Tense in the Past, Future Perfect Continuous Tense in the Past in

English. The semantics of the mentioned verb forms expresses posteriority of a result from an action or action in regard to a given orientation moment.

The verb forms in Imperfect in the Bulgarian texts of the Corpus are 710 or 13.81% of all the Bulgarian verb forms. The verb forms in Imperfect are around 1/3 of the number of verb forms in Praesens and around 1/2 of the number of verb forms in Aorist. The most frequent translation realisations of the verb forms in Imperfect in English are the verb forms in Past Simple Tense and the are more than the half of the total occurrences.

The presented data is not sufficient to draw a conclusion about the correspondence between the usage of Imperfect in Bulgarian and Past Continuous Tense in English. The highest frequency of translation realisations of Imperfect in Bulgarian have the verb forms of Past Simple Tense in English, and the translation realisations with verb forms in Past Continuous Tense are the third frequent realisation after the translation with other parts of speech.

Within the data in the Corpus there are no translation realisations in English of Imperfect in Bulgarian with verb forms in Future Continuous Tense, Future Perfect Tense, Present Perfect Continuous Tense, Future Perfect Continuous Tense, Future Continuous Tense in the Past, Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past in English. The semantics of the verb forms above and the presented statistical data lead to the conclusion that translation realisations in English with these verb forms of the verb forms of Imperfect are not probable.

The verb forms in Perfect in the Bulgarian texts of the Corpus are 269 or 5.23% of all the Bulgarian verb forms. The verb forms in Perfect are around 1/3 of the number of verb forms in Imperfect. From the presented data it can be seen that the verb forms in Perfect are with lower frequency in compare to the verb forms in Praesens, Aorist and Imperfect. The extracted statistical data from the Corpus demonstrates ambiguous behaviour in the translation realisations of Perfect from Bulgarian in English – with the largest distribution being translation realisations with Past Simple Tense in English. The second most frequent verb forms are the verb form in Present Perfect Tense, which were presented as a formal-semantic equivalent of the forms in Perfect.

Within the data in the Corpus there are no translation realisations in English of Perfect in Bulgarian with verb forms in Future Simple Tense, Future Continuous Tense, Future Perfect Tense, Future Perfect Continuous Tense, Future Continuous Tense in the Past and Future Perfect Continuous Tense in the Past in English. The presented statistical data shows that it is

unlikely that verb forms in Future tenses English forms to be used as translation realisations of the Perfect in Bulgarian. Exclusions are the verb forms in Future Simple Tense in the Past and Future Perfect Tense in the Past in English because of the fact that they are used in conditionals where they signalise for a possible terminated action.

The verb forms in Futurum in the Bulgarian texts of the Corpus are 189 or 3.68% of all the Bulgarian verb forms. The verb forms in Futurum are around 1/2 of the number of verb forms in Perfect. From the presented data for the distribution of translation realisations of the verb forms in Futurum in Bulgarian it is seen that verb forms in Future Simple Tense in English have higher frequencies and Future Simple Tense was presented as a formal-semantic equivalent of the verb forms in Futurum. The variety of distributions is significantly limited in comparison to the verb forms in Praesens, Aorist, Imperfect and Perfect.

Within the data in the Corpus there are no translation realisations in English of Futurum in Bulgarian with verb forms in Present Perfect Tense, Past Perfect Tense, Future Perfect Tense, Present Perfect Continuous Tense, Past Perfect Continuous Tense, Future Perfect Continuous Tense, Future Continuous Tense in the Past, Future Perfect Tense in the Past and Future Perfect Continuous Tense in the Past in English. The presented verb forms include all of the Perfect and Perfect Continuous tenses in English as well as the verb forms in Future Continuous Tense in the Past. The semantics of Futurum in Bulgarian expresses a posteriority of an action or a result from an action in relation to the moment of utterance. The semantics of the verb forms presented above that are not registered as translation realisations in English of the verb forms in Futurum in Bulgarian expresses an existence of a result from a terminated or continuous action in relation to a given orientation moment.

The verb forms in Plusquamperfectum in the Bulgarian texts of the Corpus are 164 or 3.19% of all the Bulgarian verb forms. The number of the verb forms in Plusquamperfectum is close to the number of verb forms in Futurum. From the presented data for the distribution of translation realisations of the verb forms in Plusquamperfectum in Bulgarian it is seen that verb forms in Past Perfect Tense in English have the highest frequencies. According to the formal-semantic description the formal-semantic equivalent of the verb forms in Plusquamperfectum in Bulgarian are the verb forms in Past Perfect Tense in English. Second highest frequency have the verb forms in Past Simple Tense.

Within the data in the Corpus there are no translation realisations in English of Plusquamperfectum in Bulgarian with verb forms in Future Simple Tense, Present Continuous

Tense, Future Continuous Tense, Future Perfect Tense, Present Perfect Continuous Tense, Future Perfect Continuous Tense, Future Continuous Tense in the Past and Future Perfect Continuous Tense in the Past in English. The presented verb forms include all of the Future tenses in English as well as the verb forms in two of the Future tenses in the Past. Given the semantics of Plusquamperfectum and the tenses listed above, it can be argued that translation realisations with verb forms in Future tenses in English of verb forms in Plusquamperfectum in Bulgarian are not possible.

The verb forms in Futurum Praeteriti in the Bulgarian texts of the Corpus are 47 or 0.91% of all the Bulgarian verb forms. The verb forms in Futurum Praeteriti are around 1/3 of the number of verb forms in Plusquamperfectum. From the presented statistical data it can be seen that the verb forms with highest frequency as translation realisations are the verb forms in Future Simple Tense in the Past in English. The verb forms in Future Simple Tense are among the verb forms with lower distribution in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" – with less than 1%.

From the presented data on the distribution of the translation realisations of the temporal forms with low frequency of occurrence in the Corpus can be concluded that they have a specific semantics, which can be transmitted through a limited number of verb forms during translation. The examples show that verb forms for Futurum Praeteriti in Bulgarian have translation realisations with verb forms in Future Simple Tense in the Past in English;

Within the data in the Corpus there are no translation realisations in English of Futurum Praeteriti in Bulgarian with verb forms in Future Simple Tense, Present Continuous Tense, Future Continuous Tense, Present Perfect Tense, Past Perfect Tense, Present Perfect Continuous Tense, Past Perfect Continuous Tense, Future Perfect Continuous Tense and Future Perfect Continuous Tense in the Past in English. The listed tenses that are not realised as translational realisations in the Corpus of Futurum Praeteriti are mainly Past and Present tenses.

Verb forms in Futurum Exactum and Futurum Exactum Praeteriti are with very low frequency within the "Bulgarian-English Parallel Corpus with Aligned Verb Forms". All of the occurrences of these verb forms are only from the examples of "Bulgarian Grammar". The verb forms in Futurum Exactum in the Bulgarian texts of the Corpus are 10 or 0.19% of all the Bulgarian verb forms and the verb forms in Futurum Exactum Praeteriti are 5 or 0.1% of all the Bulgarian verb forms in the Corpus.

The translation realisations of the forms in Futurum Exactum and Futurum Exactum Praeteriti in the examples from "Bulgarian Grammar" are only with verb forms in Future Perfect Tense and Future Perfect Tense in the Past respectively. However because of the insufficient examples the translation realisations of the verb forms in these two times cannot be adequately determined. The frequency of occurrences of verb forms in Futurum Exactum Praeteriti in Bulgarian is proof of the thesis that the complexity of the meaning restricts the distribution of Futurum Exactum Praeteriti (Nitsolova 2008: 316). The data from the Corpus can be accepted as the same proof for Futurum Exactum.

IV.5. Generalised analysis of the translation of the morphological categories from Bulgarian to English

In this part of the research statistical data extracted from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" is presented about the transfer of the meaning of the morphological categories constructing the tense hypercategory in Bulgarian and English. The translation realisation of the meanings that construct the morphological categories of tense, type of action and taxis in Bulgarian are summarized.

Statistics on the translation realisations in English of the verb forms expressing the meaning of a-synchronicity, posteriority and Anteriority in Bulgarian are presented in Table 6.

	Present Tenses in English		Past Tenses in English		Future Tenses in English		Infinitive in English		Other parts of speech in English	
	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%
A-synchronicity in Bulgarian	1532	42.79	1160	32.40	212	5.92	252	7.04	424	11.84
Posteriority in Bulgarian	28	11.24	12	4.82	179	71.89	12	4.82	18	7.23
Anteriority in Bulgarian	60	4.59	1121	85.70	7	0.54	30	2.29	90	6.88

Table 6. *Data for the translation realisations in English of the grammatical information contained by the grammatical category of tense in Bulgarian.*

The presented data shows that there are 3,580 verb forms expressing a-synchronicity in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms". The translation realisations of the meaning of a-synchronicity in regards to the morphological category of tense are various

in English – predominant distribution have the verb forms in Present and Past tenses in English. There are 250 verb forms expressing posteriority in the Bulgarian texts of the Corpus and their translation realisations are mainly with verb forms in Future tenses in English. There are 1,310 verb forms expressing anteriority and their translation realisations are mainly with verb forms in Past tenses in English. Stable behaviour of the translation realisations distinguishes forms expressing the meaning of a-synchronicity; in English posteriority is most often realised with verb forms in Future tenses, and anteriority – with verb forms in Past tenses.

The various translation realisations of the verb forms expressing a-synchronicity can be explained with the fact that in Bulgarian the meaning a-synchronicity is expressed not only by the Praesens, but also by the greater part of the "past" tenses - Imperfect, Perfect and Plusquamperfectum. The semantics of the main meaning of the "past" tenses in Bulgarian can be divided into two subtypes – relating actions or results from actions as simultaneous to a orientation moment in the past (the semantics of Imperfect and Plusquamperfectum) and relating results from actions to the moment of utterance (the semantics of Perfect). The meaning of the verb forms of the "past" tenses is transferred in English with verb forms in Past tenses (and with highest frequency of translation realisations of Perfect are verb forms in Past Simple Tense, and of Plusquamperfectum – Past Perfect Tense). From the presented statistical data a conclusion can be drawn that although the close semantics, the meaning of the morphological category of tense in Bulgarian and English have different realisations in both languages.

The statistical data for the translation realisations of the resultative and non-resultative verb forms from Bulgarian in English is presented in Table 7.

	Simple Tenses in English		Continuous Tenses in English		Perfect Tenses in English		Perfect Continuous Tenses in English		Infinitive in English		Other parts of speech in English	
	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%
Resultativity in Bulgarian	175	37.72	13	2.90	198	49.33	8	1.56	8	1.79	30	6.70
Non-resultativity in Bulgarian	3630	76.80	182	3.83	80	2.05	14	0.51	287	6.11	502	10.69

Table 7. *Statistical data for the translation realisations of the resultative and non-resultative verb forms from Bulgarian in English*

The data shows that the non-resultative forms in Bulgarian are realised primarily with verb forms in Simple tenses in English and the resultative verb forms have a various translation realisations in English with the Perfect tenses having the highest frequency. The resultative

verb forms in Bulgarian can be divided into two subtypes according to their distribution: tenses with relatively high frequency - Perfect and Plusquamperfectum; and tenses with very low frequency - Futurum Exactum and Futurum Exactum Praeteriti. From the presented data a conclusion can be drawn that non-resultative verb forms in Bulgarian are realised with verb forms in Simple tenses in English, while the resultative verb forms in Bulgarian have various translation realisations in English with Perfect tenses having the highest frequency. These various realisations can be explained with the different semantics of Perfect (simultaneity of a result with the moment of utterance); Plusquamperfectum (simultaneity of a result with an orientation moment in the past); Futurum Exactum (posteriority of a result in relation to the moment of utterance) and Futurum Exactum Praeteriti (posteriority of a result in relation to an orientation moment in the past).

The statistical data for the translation realisations of the relative and non-relative verb forms from Bulgarian in English is presented in Table 8.

	Simple Tenses in English		Continuous Tenses in English		Perfect Tenses in English		Perfect Continuous Tenses in English		Infinitive in English		Other parts of speech in English	
	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%
Relative Tenses in Bulgarian	544	59.72	73	8.32	116	12.42	7	1.19	34	3.67	136	14.69
Non-relative Tenses in Bulgarian	3261	77.40	122	2.94	162	3.68	15	0.43	260	6.17	396	9.39

Table 8. *Correlation between the translation realisations in English of the relative and non-relative tenses in Bulgarian.*

The statistical data shows that there are no prioritised distributions for the category of taxis - both relative and non-relative tenses have translation realisations with Simple tenses in English. From the data the conclusion can be drawn that there are no strictly defined translation realisations in English of the category of taxis in Bulgarian. The meanings of the relative and non-relative tenses are transferred with meanings of different morphological categories in English. This can be explained with the fact that the meaning of relativity is not limited withing a certain morphological category in the temporal system in English.

The statistical data for the translation realisations of the verb forms formed with perfective and imperfective verbs from Bulgarian in English is presented in Table 9.

	Simple Tenses forms in English		Continuous Tenses forms in English		Perfect Tenses forms in English		Perfect Continuous Tenses forms in English	
	Number of Forms	%	Number of Forms	%	Number of Forms	%	Number of Forms	%

Verb forms with perfective verbs in Bulgarian	2674	90.61%	64	2.17%	204	6.91%	9	0.30%
Verb forms with imperfective verbs in Bulgarian	1131	83.84%	131	9.71%	74	5.49%	13	0.96%

Table 9. *Correlation between translation relations in English of the perfective and imperfective verbs in Bulgarian.*

The presented statistical data shows that for the perfective verbs in Bulgarian the second most frequent translation realisations are with verb forms in Perfect tenses in English and for the imperfective verbs in Bulgarian the second most frequent translation realisations are with verb forms in Continuous tenses in English. The presented dependences might support the hypothesis for similar meaning of the perfective and imperfective verbs in Bulgarian, and the Continuous and Perfect tenses respectively in English. However the distributional frequency shows that there is no direct correlation in the meaning of the lexical-grammatical category of the verb type in Bulgarian and the meaning of the Continuous and Perfect tenses in English. Despite that it can be seen that there is a certain preference in the translation realisations of imperfective verbs with forms in Continuous tenses in English and respectively - o perfective verbs – with forms in Perfect tenses in English. Regardless of the type of the verb in Bulgarian, the verb forms in Simple tenses in English have the highest distributional frequency, which proves the hypothesis about the difference in the semantics of the categories of verb type in Bulgarian and aspect in English.

IV.6. Conclusions about the Bulgarian-English grammatical parallels with respect to the temporal forms

The formulated hypothesis that the temporal morphological categories in Bulgarian systematically predetermine the morphological categories of the corresponding translation realisations in English is confirmed according to the presented statistics: in 70% of the presented cases the temporal form in Bulgarian uniquely predetermines the temporal form in English.

The explanation of the discrepancy between the formal-semantic analysis and the empirical data from the Corpus is complex and can be considered in two directions – regarding the characteristics of the two languages and regarding the contextual usage of the verb forms. As it was discussed, in English for tense coordination implies a dependent structure of the temporal forms in the sentences and text (Rankova and Ivanova 1980: 226). In Bulgarian most of the tenses with lower frequency express background actions that serve as a temporal orientation of the main action in the sentence. The comparison between the distribution of the

45

verb forms in Bulgarian and the verb forms in English in the Corpus also gives grounds to support the hypothesis that the frequency of usage depends on the semantic characteristics of the members of a given temporal category: verb forms that express complex meanings are less common.

The temporal systems in Bulgarian and English and the morphological categories that construct them have formal-semantic similarities. The statistical data from the Corpus shows that the lexical context, the syntactic characteristics of the languages, as well as pragmatics influence the usage of the temporal forms.

V. Practical evaluation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms"

An experiment was conducted with terminology databases and software for computer-assisted translation, which aims to verify the practical application of the linguistic data and the statistical information from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms".

A part of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" has been adapted as a terminology base in order to be used as a resource in a computer-assisted translation tool: memoQ. The terminology base work on a pre-annotated text with tags for the morphological characteristics of the words which is applied to the source language in the software. Based on the tags with morphological characteristics that are present in the individual segments (sentences) from the source language (Bulgarian), the terminological base offers the most likely marker for the translation realisation of the verb form in the target language (English).

The experiment involved two translators who were familiar with the test setting, the nature of the texts, and the usage of computer-assisted translation tools. The selected texts for the experiment provide maximum variety of possible verb forms in the source language. The total volume of the texts is 1,580 word and they contain 200 verb forms in Bulgarian. The texts are divided into two groups – for testing and for evaluation, the group for testing contained 150 verb forms and the group for evaluation – 50 verb forms. After assigning the most probable tag on the target text based on the database extracted from the and the annotated source text, it was tracked in what portion of the cases the translators will either accept the assigned verb form for the verb in the current sentence or reject it.

In 85% of cases in the course of the testing, where the terminology base from Corpus assigns tag for the verb form in the target language based on the source language markers the translator agreed with the assigned verb form. In the rest of the cases the assigned verb form does not coincide with the choice of the translator: in 9% of those cases the verb form chosen by the translator coincides with the second most likely assigned form and in the other 6% – with another consecutive form. There is no case in which the translator has chosen a verb form that was not from the proposed list of possible verb forms.

The result from the experiment for the practical application of the Corpus is that the verb forms in the two languages can be automatically correlated based on the alignment models

contained in the Corpus, and enriched translation memories and models with linguistic information can be used both to assist the translation process and to automatically edit already translated texts.

VI. Conclusion

The research offers a theoretical approach for analysing the characteristics and the organization of the temporal verb systems in Bulgarian and English. The approach of the research was chosen because it has been proven as successful in comparative studies of languages with different structures. The approach of the research is based on a comparative examination of the formal-semantic description of the temporal categories in the two languages and a statistical model for the translation of the temporal forms extracted from a specifically designed parallel bilingual corpus: "Bulgarian-English Parallel Corpus with Aligned Verb Forms"

VI.1. Summary

The presented research is relevant to the increasing interest in the scientific field of various applied scientific developments and researches for description and analysis of linguistic phenomena through the methods of computational linguistics and in particular through the methods of the corpus-based approach and statistical modelling. The purpose of the research is a comprehensive and consistent description of the transfer of grammatical information expressed by the verb forms in translation from Bulgarian to English. The research on the interaction between the temporal systems in Bulgarian and English and the morphological categories that construct them has theoretical value for clarifying the patterns of correlation between the grammatical information, the semantic structure and the means of expressing the temporal relations in both languages.

The analysis of the grammatical parallels of the data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" provides: an overview of the data for the Bulgarian and the English temporal forms separately; an overview of the data for the transfer of information of the morphological categories of tense, type of action and taxis in Bulgarian to the morphological categories of tense and aspect in English.

The presented statistical data indicate that the theoretical hypothesis of conformity in the formal-semantic analysis of the two languages can be proven through the chosen approach of the research, and the statistical data also demonstrates that in many cases the temporal form in English as a target language depends on the combinatorial presence of certain morphological categories in Bulgarian as a source language. The importance of the context for the translation of verb forms from Bulgarian to English is an obstacle for complete confirmation of the

hypothesis: in the formal-semantic analysis the context is not taken into account, while in the analysis of the empirical cases of the translation realisations registered in the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" the context plays an essential role in forming the temporal meanings.

The presented statistical data considered independently from the theoretical hypothesis, shows that the verb forms with the highest frequency of distribution have relatively diverse translation realisations, whereas the verb forms with the lower frequency of distribution have a relatively limited possible translation realisations in English. Also certain temporal forms are not registered in the language data of the Corpus, which serves as a proof of the thesis for their limited usage. The following conclusions can be drawn from the research:

1. The temporal systems of Bulgarian and English have sufficient similarities in formal-semantic terms, which is a prerequisite for a comparative research and makes possible the construction of a reliable translation model based on the semantic properties of the morphological categories that construct the temporal systems in both languages.

2. The morphological categories that construct the temporal systems in Bulgarian and English have similar formal-semantic characteristics and suggest similar modelling of the grammaticalised meanings in the two languages.

3. The presented statistical data extracted from the Corpus proves that the combinatorial presence of certain morphological categories in Bulgarian predetermines the temporal form in English during translation. The means of expressing the grammatical information are different between the two languages. The presented statistical data also demonstrates the trend that the most common verb forms in Bulgarian have relatively diverse translation realisations in English, while the more complex and less common verb forms in Bulgarian have relatively restricted translation realisations.

The conclusions from the scientific approach chosen for the analysis of the object of the research are:

1. The corpus-based approach is functional and successful in examining the grammatical parallels between the temporal systems in Bulgarian and English.

2. The volume and the representativeness of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" are sufficient to answer the formulated questions of

the research, but they cannot provide a complete comprehensive description of the characteristics of the temporal systems in Bulgarian and English. The low distributional frequency and the specific usage of some temporal forms are the reason for the insufficiency of reliable linguistic data for a comprehensive analysis of them.

3.   The conducted experiment with a terminology base containing grammatical information for the translation realisations from the Corpus implemented in computer-assisted translation software proves the successful application of parallel language models for translation improvement. The data from the Corpus can be applied as a supplementary resource for computer aided translation systems, as well as in the selection of monolingual and parallel training data for machine translation systems.

The research presents a relevant theoretical and applied problem related to a complex language phenomenon. The problem is presented by a modern and scientifically proven research approach, through which the objectives of the study are realised and empirical and quantitative data are provided for the object of study: the Bulgarian-English grammatical parallels in the translation of verb forms from Bulgarian to English.

VI.2.   Contributions of the research

The scientific contributions of the research can be summarized in several main areas, which correspond to the organization of the parts of the dissertation, and can be formulated as follows:

Formulating an unified theoretical description of the temporal systems in Bulgarian and English both individually and in comparison. Presenting the development of the theoretical views about the categories that construct the temporal meanings of the verb forms in Bulgarian and English. The mechanism of constructing the grammatical temporal meaning in the verb forms in Bulgarian and English is examined in a comparative plan and the basic mechanisms in the interaction between the temporal morphological categories in the two languages are formulated. A formal-semantic comparison between the meanings of the temporal categories in Bulgarian and English is presented.

Construction and creation of the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" During the research no information was found on the availability of another parallel resource for Bulgarian and second Language in regard to the verb forms. The

process of resource selection, pre-processing, annotation and software applications used in the working process on Corpus are presented.

The "Bulgarian-English Parallel Corpus with Aligned Verb Forms" provides opportunities for exploring the grammatical parallels between Bulgarian and English. The Corpus can be considered as a resource for a formal description of the grammatical parallels between the temporal systems of Bulgarian and English and a source for extracting statistical models for the relations of the translational realisations of the verb forms in both languages. The linguistic data from the Corpus can be used to enrich translation models for the purposes of machine translation as well as for theoretical studies. The presented working process of the research in regard to the annotation and the structure of the Corpus can serve as a guide for development of similar linguistic resources for different researches of similar linguistic phenomena.

Objective quantitative and distributive information about verb systems in Bulgarian and English is presented. The information is suitable for different types of scientific analysis outside the scope of this research. Credible linguistic data to support the creation of various language applications for computer processing of the natural language are provided – such as a statistical annotation model and a statistical translation model for verb form.

A supplementary contribution of the research is the integration of the linguistic data from the "Bulgarian-English Parallel Corpus with Aligned Verb Forms" in a computer-assisted translation software for its practical application as a resource. In this way the statistical data can be used in the translation process and can be corrected and enriched with additional language material.

Literature

Aleksova 2003: Aleksova, Kr. Sintagmatitshnata neutralizacia na gramatitsheskite opozicii kato pritshina za razlitshnite sastoyania, koito morfologitshnata sistema mozhe da zaema (warhu material ot temporalnata sistema na balgarskia ezik). – In: *Materiali ot Mezhdunarodnia slavistitshen simpozium v ramkite na Yubileynite Paisievi tshetenia.* Plovdiv: Plovdivski universitet. [http://belb.info/personal/aleksova/neutralizacii_1.pdf] (accessed on 20.02.2020)

Andreychin 1938: Andreychin, L. Obshta harakteristika na balgarskata glagolna sistema (translation of Ogólna charakterystyka bułgarskiego system koniugacyinego, 1938). – In: Pashov, P. Nitsolova, P. (eds.) *Pomagalo po balgarska morfologia. Glagol.* Sofia: Nauka i izkustvo.

Andreychin 1944: Andreychin, L. *Osnovna balgarska gramatika.* Sofia: Nauka i izkustvo.

Atanasova et al. 1963: Atanasova, T. Rankova, M. Rusev, R. Radulova, N. *Angliyska gramatika v sravnenie s balgarski ezik.* Sofia: Nauka i izkustvo.

Baskervill and Sewell 1895: Baskervill , W. Sewell, J. *An English Grammar: For the Use of High School, Academy, and College Classes.* New York: American Book Compnay [https://archive.org/stream/anenglishgramma00sewegoog#page/n156/mode/2up] (accessed on 20.02.2020).

Valeika and Buitkiene 2003: Valeika, L. Buitkiene, J. *An introductory course in theoretical English grammar*. Vilnius: Vilnius Pedagogical University.

Veiga 1991: Veiga, A. *Condicionales, concesivas y modo verbal en español.* Santiago de Compostela: Universidade de Santiago de Compostela.

Garavalova 2003: Garavalova, Il. *Shkolata na naturalnata morfologia i balgarskata morfologicheska sistema.* Sofia: SemaRSh

Gerdzhikov 1984: Gerdzhikov, G. *Preizkazvaneto na glagolnoto deystvie v balgarskia ezik.* Sofia: Nauka i izkustvo.

Gerdzhikov 2000: Gerdzikov, G. Kategoriyata vreme kato hiperkategoria – In: *Balgarski ezik*, No. 1, 2000 [https://liternet.bg/publish/ggerdzhikov/hyper.htm] (accessed on 20.02.2020).

Johnson 1640: Johnson, B. *The English Grammar*. 1640, London [https://archive.org/details/englishgrammar00jons/page/n8] (accessed on 20.02.2020).

Kabakchiev 2003: Kabakchiev, K. *Angliyska gramatika.* Sofia: Pensoft.

Klein 1994: Klein, W. *Time in Language*. London: Routledge.

Koeva 2014: Koeva, Sv. Balgarskiyat natsionalen korpus v konteksta na svetovnata teoria i praktika. – In: *Ezikovi resursi i tehnologii za balgarski ezik.* Sofia: Akademitshno izdatelstvo "Prof. Marin Drinov", 29 – 53.

Koehn 2010: Koehn, P. *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Comrie 1985: Comrie, B. *Tense*. Cambridge: Cambridge University Press.

Kutsarov 2007: Kutsarov, Iv. *Teoretitshna gramatika na algarskia ezik: morfologia.* Plovdiv: Universitetsko izdatelstvo "Paisii Hilendarski"

Leech 1991 Leech, G. The state of the art in corpus linguistics. – In: *English Corpus Linguistics*, London: Longman, 8 – 29.

Lewis and Mol 2013: Lewis, G. Mol, H. *Grammar for Young Learners*. Ofxord: Ofxord University Press .

Manning et al. 2013: Manning, L. Cassel, D. Cassel, J. C. St. Augustine's Reflections on Memory and Time and the Current Concept of Subjective Time in Mental Time Travel. – In: *Behavioral Sciences*, 2013, 3, 232 – 242.

Murphy 1991: Murphy, R. *English grammar in use*. Cambridge: Cambridge University Press.

Nitsolova 2008: Nitsolova 2008, R. *Balgarska gramatika: Morfologia.* Sofia: Universitetsko izdatelstvo "Sv. Kliment Ohridski".

Nitsolova 2017:Nitsolova 2017: Nicolova, R. *Bulgarian Grammar*. Berlin: Frank & Timme GmbH.

Paskaleva 1987: Paskaleva, E. *Kompyutar i prevod.* Sofia: Nauka i izkustvo.

Reichenbach 1947: Reichenbach, H. *Elements of Symbolic Logic*. New York: The Macmillan Company, 1947, 287-298.

Rankova and Ivanova 1980: Rankova, M. Ivanova, Tsv. *Angliyska gramatika.* Sofia: Nauka i izkustvo.

Stambolieva 2015: Stambolieva, M. Klasove dumi: angliysko-balgarski paraleli. – In: *Godishnik na departament "Anglicistika".* Sofia: NBU.

Lowth 1762: Lowth, R. *A Short Introduction to English Grammar with Critical Notes*. Philadelphia: Aitken [https://archive.org/stream/shortintroductio00lowtrich#page/34/mode/2up] (accessed on 20.02.2020).

Filipova 2002: Filipova, M. *Praktitsheska angliyska gramatika.* Sofia: EMAS.

Firat et al. 2016: Firat, O. Cho, K. Bengio, Y. Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism. – In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 866–875*

House and Stevens 2000: House, C.; Stevens, J. *Grammar no problem*, 2000, Berlin: Cornelsen.

Hornstein 1990: Hornstein, N. *As Time Goes By: Tense and Universal Grammar*, Cambridge: MIT Press.

Hutchins 1995: Hutchins, W. Machine Translation: A Brief History. – In: Koerner, E. Asher, R. (eds.) *Concise history of the language sciences: from the Sumerians to the cognitivists.* Oxford: Pergamon Press, 431-445.

Tsergova 2012: Tsergova, V. *Konyunktivniyat imperfekt v savremennia portugalski ezik.* Sofia: Universitetsko izdatelstvo "Sv. Kliment Ohridski".

Tsergova 2014: Tsergova, V., *Yerarhitshnost na kategorialnite znatshenia i formi. – In:* Krastev, N. (ed.) *Mislene, ezik, retsh.* Sofia: Universitetsko izdatelstvo "Kliment Ohridski".

List of author`s publications on the research

Lazarov 2016a: Lazarov, T. Osobenosti na glagolnite sistemi i natshinite za izrazyavane na vremeto v balgarski i angliyski. Semantitshen transfer pri prevod na glagolnite formi ot balgarski na angliyski. – In: *„Littera et Lingua"*, No. 1, 2016.

Lazarov 2016b: Lazarov, T. A Possible Solution to the Problem of Machine Translation of Verb Forms from Bulgarian to English. – In: *Proceedings of the Second International Conference Computational Linguistics in Bulgaria (CLIB 2016), 95 – 100.*

Lazarov 2016c: Lazarov, T. Analiz na resursite za statistitsheski model za prevod na glagolnite formi ot balgarski na angliyski. – In: *Balgarski ezik, No. 63 (2016), 4* 95 – 101.

Lazarov 2017: Lazarov, T. Funktsionalni gramatitshni paraleli pri prevoda na glagolnite formi ot balgarski na angliyski. – In: *Dokladi ot Mezhdunarodnata yubileyna konferencia na Instituta za balgarski ezik "Prof.* Lyubomir Andreytshin (15/16 May 2017)", 312 – 327.

Lazarov 2018: Lazarov, T. Bulgarian-English Parallel Corpus for the Purposes of Creating Statistical Translation Model of the Verb Forms. General Conception, Structure, Resources and Annotation – In: *Proceedings of the Third International Conference Computational Linguistics in Bulgaria (CLIB 2018).* 193 – 202.

Lazarov 2019a: Lazarov, T. Balgaro-angliyski paralelen korpus sas saotneseni glagolni formi. – In: *Sapostavitelno ezikoznanie, XLIV (2019), 1.* 5 – 14.

Lazarov 2019b: Lazarov, T. Savramenno sastoyanie na Balgarsko-angliyskia paralelen korpus sas saotneseni glagolni formi. – In: *Balgarski ezik, No. 1, 2019, 103 – 113.*