FRANTIŠEK MARTÍNEK
Institute of Czech Language and Theory of Communication
Charles University of Prague, Faculty of Arts,
KATEŘINA RYSOVÁ
Institute of Formal and Applied Linguistics
Charles University of Prague, Faculty of Mathematics and Physics,

A CORPUS OF OLDER CZECH TEXTS FOR LINGUISTIC PURPOSES[1]

SŁOWA KLUCZOWE: korpus diachroniczny, język czeski XVI wieku, transliteracja, szyk wyrazów, swobodne modyfikacje wyrazowe

KEYWORDS: diachronic corpus, humanistic Czech, transliteration, word order, free verbal modifications

1. Introduction

In the first part of the paper, the principles of the Corpus of Humanistic Czech[2] are presented, with the focus on the linguistic aspects of the building of the corpus and searching within the corpus. The selected method of transcription is discussed and the theoretical reasons for it are given. Furthermore, the selection of the texts included is explained.

In the second part, the results of an analysis of word order in humanistic Czech, based on the Corpus of Humanistic Czech, are presented, and in this

---

[2]   The adjective *humanistic* does not refer to any special text types but to the first period of Middle Czech; the second one is called Baroque Czech (1620–1780).

way, exploration possibilities of this corpus are shown, with regard to syntactic phenomena. In particular, we explored the position of obligatory verbal modifications, in the sense of them being verb-dependent participants, which are dependent on verbs of the type *zacházeti*, *nakládati s něčím* ('to deal with something').

2. Corpus of Humanistic Czech

2.1. General principles

The Corpus of Humanistic Czech was created as the material base of the *Linguistic Analysis of Czech Humanistic Texts* project in 2011. It will be included in the diachronic part of the Czech National Corpus as a guest corpus.

It is a balanced (see 2.2) electronic corpus of texts printed from 1500 to 1620. It includes more than half a million word forms and consists of approximately 50 texts and extracts of longer texts, respectively. In March 2011, all the texts were digitised (by students) and approximately two thirds have been already checked. The purpose of the balanced corpus is first of all didactic: it is intended for morphological, syntactical and lexical analyses for use on university bachelors' and masters' courses of the Czech language.

The texts included are transcribed, i.e. transformed into the Modern Czech orthographical system with respect to phonetical and phonological peculiarities of the older language.

Any problematic or questionable phenomena in texts, where the script enables two (or more) interpretations, are recorded in transliteration. The main reason for this method is that transliteration does not enable any effective searching in electronic texts (cf. Kučera 1998: 306f.). Moreover, the Czech 16th century orthography – the so-called *bratrský pravopis*, "Czech Brethren orthography" – is heterogeneous, and using transliteration, one has to solve analogous problems as with transcription.

Annotation tags are embedded in the text, under similar principles which are used in the Czech diachronic corpus DIAKORP. The only difference is caused by the fact that lemmatization of the corpus is not planned. Irregular word forms are thus marked and complemented by the default form, which enables the user to search through it easily and more effectively. Thus, these default forms called *lemmas* are an auxiliary instrument.[3]

---

[3]   But one has to keep in mind that this approach makes the potential total lemmatization much easier, if the texts were to be used in future research.

Due to technical issues connected with access to the corpus, the resources included are plain text documents; also one can access them using common searching programmes like MonoConc. Restricted time and scarce number of collaborators made it impossible to invent and apply a more sophisticated kind of text processing, e.g. multi-level html-annotation.[4]

The transcribed text looks as follows:

> *<a>Abraham z Gynterrodu*
> *<t>Cyri paedia</t>*
> *<r>1605</r>*
>
> *<s>1a</s><k><o>Cyri paedia</o>, hodnověrná starožitná historia o chvalitebném ve všelikých <e>wewsselikých</e> knížecích <e>Knjžetcých</e> ctnostech vycvičení a zvedení, o slavných skutcích, vítězných válkách a právě heroitském <l>heroický</l> šlechetném životu Cýra Staršího, prvního monarchy perského.*
> *V níž se vypisuje, jakým spůsobem <l>způsob</l> Pán Bůh monarchí <l>monarchie</l> kaldejskou <l>chaldejský</l> skrze Daria […] a Cýra […] na médský <e>Medský</e> a perský národ přenesti <l>přenésti</l> a monarchí médskou <e>Medskou</e> a perskou zřídit i a utvrditi ráčil. […] </k>*
> *Léta Páně 1605. <e>M. DC. V.</e>*
>
> (XenCyr1605 1a)

As hinted above, the basic logical as well as physical structure of the texts (chapters and their titles, paragraphs, pagination, marginal notes etc.) is reflected in the transcriptions. The symbols *<a>*, *<t>* and *<r>* inform about the author (or the translator, respectively), title and year of the print's origin. The symbol *<s>* indicates the foliation; the *<k>* the name of paragraph titles etc., here the title of the print; the *<e>* transliterations of questionable words or word groups. The *<l>* indicates the lemma (meant to enable elementary searching in the corpus, see above). The symbol *<o>* marks the so-called *quotations*: these are foreign words in the Czech text which are not integrated into the Czech language, as far as their phonological and morphological properties are concerned.

In addition to this balanced corpus, a larger corpus of two million word forms was prepared for verifying results and hypotheses. Its task was to include as many texts from 1500–1620 as possible. Due to the fact that quantity is important here, the data in this corpus could not be either controlled manually or provided with annotation tags.

---

[4]   For more information see Meyer 2005. In his paper, a specific Slavic diachronic corpus – the Regensburger Russian one – is taken into account.

2.2. Text types in the corpus[5]

Let us show how the texts which were included in the balanced corpus have been selected. The original idea to create a representative corpus had to be re--thought. There were problems with text types definitions. When we decided to use the eight text types delimitated by Zikánová (2008: 51), we had to include very short parts of texts into the corpus. Then, the lexical items included would be far from representative because the text parts included would be random.

For classification of text types, one may be inspired by other historical corpora of individual languages (for German, cf. Bergmann, ed., 1998). Furthermore, research is carried out on very different text types. According to the definition of such types, individual texts are chosen for corpora, mainly typical ones with as huge an impact as possible (see ibid.: 24, 32, 43, 72 etc.). Texts from the period examined may be sorted proportionally with respect to their date of origin, but it is also possible to stress some parts of the period or collect texts written in certain time distances to highlight differences between them (cf. ibid.: 42).

Moreover, the following questions appeared: how can the reception of the texts be measured, and should it be taken into consideration? Can texts from as many printers as possible be included? How should the division into Catholic and Evangelic authors be reflected?

After these considerations, the decision was taken to create a balanced corpus, and a larger corpus as its supplement. Thus, in the balanced corpus, texts of various types and dates of origin have to be represented. Texts from the whole period (with a focus on the Veleslavín era) equally represented as to time and text types have been included. All text types described in Zikánová (2008: 51) have been included. With respect to the planned relatively small size of the corpus, further examination and specifications (i.e. solutions of the questions above) were not needed.

2.3. Theoretical background of the transcription

With respect to the aims of the corpus, many graphical and orthographical phenomena have been changed without being noted in the transcription, e.g. abbreviations which are common in prints, graphical variants of some letters (two letters for *s*, *r* or *b*), the distribution of capital and small letters, the use of Gothic and Latin type. Many orthographical variants are unified by the transcription; therefore they cannot be identified retrospectively: *wijera*/*wiera* → *viera*, but *wijra*/*wjra* → *víra*, *naylepſſij*/*naglepſſij* → *najlepší*. The original punctuation has been changed to the syntactic one.

---

[5]   About the text types included in the corpus see Martínek 2009: 461 a Martínek – Ocelák 2010: 220f.

Another task is to note the difference between peripheral and central, systematic phenomena. (As the peripheral ones concern primarily phonetics, and our assessment may be influenced by the graphics of the prints and printers' mistakes, a connected problem is the difference between (ortho)graphical and language phenomena; see below). The peripheral phenomena are kept in the transliteration only, not in the transcription (see also 1.4, point 2).

The theoretical principles for dealing with questionable phenomena are: determination of the default forms of the lexical units is theoretically grounded in the three-level approach to the lexical and morphological phenomena. The theory of German linguist L. Lemnitzer – which advocated for the differentiation of tokens, words, and lemmata in synchronic corpora – was modified, and the phenomena were classified into individual, collective, and systematic.

What are the criteria that note the difference between these three groups? Let us explain this standpoint using several examples of vowel quantity in roots and declension endings.

  a. An **individual anomaly** will be emended in the transcription but recorded in the transliteration of the word. Different irregularities fall here, e.g. irregular vowel quantity in declension endings or in roots where only the printers' imperfection can be presumed etc.

  b. A **collective special feature (distinction)** is inherent to a group of authors or texts. It is marked in transcription; a lemma may be added where appropriate. Examples from vowel quantity: *litost* in place of *lítost*; *míle*, *kníha*, *pozdvíhnouti, náděje* etc. in place of *-i-/-a-*, probably also in imperatives of verbs of the 6th infinitive class (with the infinitive suffix *-ova-* and the present suffix *-uj-*) like *nekupůj/nekupuoj* in place of *nekupuj* and doubtless in masculine nominatives like *muží, doktoří* in place of *muži, doktoři*.

  c. The description of **systematic phenomena** would take place by lemmatizing the corpus and its following lexicographical processing. As examples of systematic phenomena which need not be evident at first sight, let us name vowel quantity unification in neutral substantives like *kázaní* (where the switch to *kázání* takes place just in the 16th century) and and the changing of the root vowel in infinitives when a syllable prefix is added (*dáti – podati*, but also *bíti – zabiti* and analogical, i.e. not system-based, *vésti – přivesti*).

The juxtaposition of various texts enables us both to include the phenomena in these three groups and to give reasons for this classification. Electronic access to the texts enables an easy revision of preliminary conclusions.

When editing the texts, the following objection to our approach may appear: at first sight, it may seem that making changes in the text, here by unifying some "individual" phenomena (e.g. vowel quantity), could mean levelling and reduction of information. However, processing the text using computers gives

the possibility to process the text at more levels (as mentioned above). Therefore, the problem is not **whether** one should note/register e.g. an anomalous marked vowel quantity, as it is rather evident that we should. This problem is **on which level of treating the text** it should be recorded.

2.4. Expected contributions

The expected contributions of the corpus are as follows:

1. We hope that this material-based work with texts enables a more precise description of many grammatical and lexical phenomena of 16th century Czech. Older descriptions usually tend to stress unusual and unexpected phenomena, and they only compare an older language stage either with today's one or with a linguist's personal opinion about the older stage.

2. The common distinction between (ortho)graphical and language phenomena, which is used by most Czech editors should be used more exactly for dealing with older Czech texts. It is clear that an editor should change only the orthographical phenomena, not language ones, while transcribing a text. Yet there appears to be an area of transitory phenomena which stand between orthography and language. This is an important reason for using the parallel transliteration of words or word groups which can be interpreted ambiguously. To be specific, the status of many transitory phenomena will be possible to define after elaborating and describing a much more extensive text material than the announced half million of words. However, it is possible to verify some hypotheses at present. For example, there may be arguments for affection of the language system by the (ortho)graphical usage.

3. There is also a methodological contribution. While theoretical problems of the transcription have been solved, there also appeared two general instructions: First, how to select one possible solution from a set which seem to be equally good, and second, how to create a boundary in a fuzzy set, in a continuum of phenomena.

4. The last but not least point is the cooperation with the diachronic part of the Czech National Corpus (Diakorp) and with the Institute for Czech Language of the Czech Academy of Sciences. Our project enables us to include our texts in the Diakorp as well as enlarge its material base significantly, as far as Humanistic Czech is concerned, and thus to increase the possibilities of study.

3. Usage of the corpus for a word-order research

Texts from the Corpus of Humanistic Czech may be used in many ways. Similar corpora are typically used for lexicographical and lexicological research

but the system of processing this corpus also allows us to examine the linguistic phenomena of higher levels of language. The results presented here consider a small word-order-research based on corpus texts of the period 1500–1620. Attention will be focused on the placement of so-called free verbal modifications expressing manner in competition with so-called actants. This means it will deal with sentences like:

> [Ty jsi; K.R.] přístup sobě k dosáhnutí toho království strojil, <velmi oulisně a přívětivě s knížaty a lidem země francské zacházel>.
> < … (you have) <u>very greasily and amiably</u> <u>with princes and people of Frankish country dealt></u>[6]

The participants like *velmi oulisně a přívětivě* ('very greasily and amiably') are marked as free verbal modifications (in many respects, they are identical with adverbials); the participants like *s knížaty a lidem země francské* ('with princes and people of Frankish country') are marked as actants (inner participants).

In the research, the basic word order of free modifications expressing manner and the actants (namely patient) were examined. We examined only sentences with certain characteristics. The research was based on the affirmative declarative sentences with the predicative verb from a group of verbs with three obligatory participants, i.e. participants which are necessary for the grammatical and especially for the semantic completeness of the sentence. The first obligatory participant was the actor, the second was the patient and the third was the obligatory modification expressing manner: the valency frame of them was ACT (obl.), PAT (obl.), MANN (obl.). The research was focused on the sentences with the verbs that semantically express mental action – *zacházeti* ('deal'; valency frame in Czech: somebody, with someone/something, somehow) and *nakládati* ('dispose'; valency frame in Czech: somebody, with someone/something, somehow). On the basis of sentences from the corpus, we investigated whether it is more typical to use "with someone somehow" or "somehow with someone" in older Czech.

3.1. Results

In the corpus material, there were about 40 sentences that met the established conditions. Most of them were sentences with the verb *zacházeti*. The material from the corpus is not large and therefore the results cannot be generalized but some word-order tendencies in older Czech were indicated by them.

The analysis of the sentences showed two major word-order tendencies:

---

[6]   The translations of Czech sentences given as examples follow Czech word order.

1. The basic word-order-position of a context-unbound obligatory patient (PAT) and a context-unbound obligatory free modification expressing manner (MANN) in Czech 1500–1620:

MANN – PAT

Example 1:
A v městě Brodě Uherském tolikéž vnově najatý soldát učiněn byl profousem (dosti postavy krásné a spanilé) a tolikéž <<u>nekřesťansky</u> <u>s lidem obecním</u> zacházel>.

< … (he) <u>unchristianly</u> <u>with ordinary people</u> dealt>

Example 2:
Když sme pak k Solnaku přitáhli, tu sme novinu uslyšeli, kterak jest Hatvan od našich vzat, <<u>jak pesky a nekřesťansky</u> Balouni <u>s Turky, s ženami i s dětmi jejich</u> sou zacházeli, těhotné ženy rozřezovaly, dítky při prsech napoly rozpoltili>.

< … <u>how like with dogs and unchristianly</u> "*Balons*" <u>with the Turks, with women and with their children</u> dealt … >

Example 3:
I dí: Nebude více nazýváno jméno tvé toliko Jákob, ale také Izrael; <nebo jsi <u>statečně</u> zacházel <u>s Bohem i lidmi></u>, a přemohls.

< … because you have <u>bravely</u> dealt <u>with God and people</u> … >

2. The basic word-order position of context-bound obligatory patient (PAT) and context-unbound obligatory free modification expressing manner (MANN) in Czech 1500–1620:

PAT – MANN

Example 4:
Proti tomu neučiní-li to, že jim budou živnosti zastavené, vojáky že mušejí chovati, kteréž že je bíti <a <u>s nimi</u> <u>co nejhůř</u> zacházeti že budou>.

< … and that <u>with them</u> <u>as badly as possible</u> (they) will deal>

Example 5:
V nenávisti máte dobré a milujete zlé, násilně snímáte kůži z lidu mého i maso z těla jejich, jedli ste tělo lidu mého a z kůže jste je vyvlékli a kosti jejich jste zlámali a jako v moždíři stloukli <a zacházeli ste <u>s nimi</u> <u>jako s masem uprostřed hrnce></u>.

< … and you dealt <u>with them</u> <u>like with meat in the middle of the pot</u>>

Example 6:
V čemž naříkáno bylo na některé přední pány a biskupy uherské, <že by <u>s týmž králem a pánem jich mladým</u> <u>neupřímně a nevěrně podle vůle své</u> zacházeli, málo ho sobě vážíce>.

< … that they would <u>with the same king and their young lord</u> insincerely and unfaithfully

<u>according to their will</u> deal … >

It was shown that contextual boundedness of patient strongly influences the basic sequence of the participants in the structures examined. It is interesting that in our material, there were patients more often contextually bound than free modifications expressing manner. It is probably more typical to express first with whom it is dealt and then, in the following sentence, to add the following piece of information, and how to deal with him or her.

In the material from the corpus, there were also sentences that did not follow the major word-order tendencies introduced above. The most typical word-order pattern of these 'exceptions' was:

contextually unbound MANN – contextually bound PAT

In some cases, this order was caused by syntactic needs, for example, in the sentences with coordination:

Example 7:
<*Žena <u>lépejí</u> jedním prstem dítek se dotýkati a <u>s nimi</u> zacházeti umí nežli muž všemi oudy těla.*>

<Woman <u>better</u> with one finger touch the children and <u>with them</u> deal can than man with all members of the body.>

However, in the material, there were also other sentences with the word-order pattern of 'exceptions'. This means that this word-order-pattern in older Czech is also possible but not typical.

Example 8:
I byl nákulhavý na obě nohy, na obě strany, a tu se na Horách Kutnách osadil, pojav k manželství Alžbětu, dceru nebožtíka kněze Jakuba Melisea, děkana tu na Horách Kutnách, <a <u>nesvorně</u> <u>s sebou</u> zacházeli>.

<and (they) <u>unconcordantly</u> <u>with themselves</u> dealt>

3.2. Juxtaposition with modern Czech

The findings based on the older Czech material are the same as the findings for contemporary Czech. The Czech linguists E. Hajičová, P. Sgall and E. Buráňová (1980) have conducted a study in Czech word order in a similar manner (but not using a corpus). They examined the contextually unbound part of the Czech sentence and studied whether there was a tendency of the sentence components to create the sentence using some word order patterns. They discovered that there

was quite a strong tendency of contextually unbound items to stand in sentences in a certain order (in so-called systemic ordering). According to E. Hajičová, P. Sgall and E. Buráňová (1980: 77), in systemic ordering of Czech, MANN comes first and is followed by PAT. However, in their research, there was no respect for individual semantic and syntactic features of verbs. The study was based on sentences with various predicative verbs (they had various lexical meanings and also valency properties). Their results were thus verified with the use of the contemporary Czech material from the Prague Dependency Treebank, the Czech National Corpus and from texts on the Internet. We have searched only for sentences with the predicative verb *zacházet* and *nakládat* and we have checked the word order of items expressing MANN and PAT which were contextually unbound. This small exploration of about 40 sentences has confirmed that in contemporary Czech sentences with the predicate *zacházet* and *nakládat* there is a tendency (bud no fixed rule) to put MANN first and PAT next:

> Example 9:
> <*Rada Evropy: Slovensko špatně zachází s vězni.*>

> <Council of Europe: Slovakia badly deals with prisoners.>

> Example 10:
> <*Muž neopatrně zacházel se zbraní,*> prostřelil si hlavu.

> <A man carelessly manipulated with a gun…>

> Example 11:
> <*Mají- li tedy naše ministerstva pravomoc svobodně nakládat s prostředky, jež jim parlament vyčleňuje ze státního rozpočtu,*> musejí podle Páva odpovídat sama i za výběr podniků.

> <So if our ministries have the power to freely dispose of funds which are allocated them by the Parliament from the state budget…>.

Abbreviations:

MünstKozm1554 – Muenster S. 1554: *Kozmografia česká*. Translated by Zikmund from Púchov, Prague. Knihopis' No. 5969.

XenCyr1605 – Xenofon 1605: *Cyri Paedia. Hodnověrná starožitná historia […] o slavných skutcích […] Cýra staršího*. Translated by Abraham from Gynterrod, Prague. Knihopis' No. 17061.

References:

Probleme der Textauswahl für einen elektronischen Thesaurus. Beiträge zum ersten Göttinger Arbeitsgespräch zur historischen deutschen Wortforschung 1. und 2. November 1996, ed. R. Bergmann, Stuttgart – Leipzig 1998.

Český národní korpus – Czech National Corpus [online]. [2010–05–12]. Institute of the Czech National Corpus, Faculty of Arts, Charles University, Prague, http://www.korpus.cz.

Daneš F., 1971, *Větné členy obligatorní, potenciální a fakultativní*, [in:] "Miscellanea Linguistica" 1971, pp. 131–138.

Gebauer J., 2007 (reprint): *Historická mluvnice jazyka českého. Díl IV, Skladba*, Prague.

Jungmann J., 1989 (1st issue 1835–1839), *Slovník česko-německý I–V*, Prague. Also available at www.slownjk.cz.

Kučera K., 1998, *Diachronní složka Českého národního korpusu: Obecné zásady, kontext a současný stav*, in: "Listy filologické" 121, pp. 303–313.

Lemnitzer L., 1997, *Akquisition komplexer Lexeme aus Textkorpora*, Tübingen.

Manuscriptorium [online] [2010–10–12]. Czech National Library, www.manuscriptorium.com.

Lopatková M., Žabokrtský Z., Kettnerová V. et al., 2008, *Valenční slovník českých sloves*, Prague.

Martínek F., 2009, *Korpus českých textů z období humanismu jako východisko lingvistických analýz*, in: *Varia XVIII. Zborník príspevkov z XVIII. kolokvia mladých jazykovedcov*, Prešov, pp. 457–464 (electronic issue).

Martínek F., Ocelák R., 2010, *Tvorba korpusu k lingvistické analýze humanistické češtiny*, [in:] *Varia XIX. Zborník príspevkov z XIX. kolokvia mladých jazykovedcov*, Trnava, pp. 219–227 (electronic issue).

Meyer R., 2005, *The Regensburg Diachronic Corpus of Russian: A New Source for Linguistic Research (Not Only) on Modality*, [in:] *Modality in Slavonic Languages. New Perspectives*, eds. B. Hansen, P. Kralik, Munich, pp. 315–336.

Mikulová M. et al., 2005, *Anotace na tektogramatické rovině Pražského závislostního korpusu: anotátorská příručka*, Prague.

Nejedlý P. et al. 2010, *Lexikální databáze humanistické a barokní češtiny [Lexical Databank of Humanistic and Baroque Czech]* [online] [2010–10–12]. Institute for Czech Language CAS, Prague, http://madla.ujc.cas.cz.

Panevová J., 1980, *Formy a funkce ve stavbě české věty*, Prague.

Pražský závislostní korpus – Prague Dependency Treebank. Version 2.0 [2010–05–07]. Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, http://ufal.mff.cuni.cz.

Rysová K., 2011, *K postavení obligatorních směrových doplnění "odkud" ve větě češtiny doby střední*, [in:] "Bohemica Olomucensia" 2011,

Rysová K., 2009, *Vliv valence predikátového slovesa na slovosledné uspořádání adverbiálií*. Unpublished thesis. Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University, Prague.

Sgall P., Hajičová E., Buráňová E., 1980, *Aktuální členění věty v češtině*, Prague.
Vokabulář webový [“Web-Vocabulary“] [online]. Version 0.4.2. [2010–06–05]. Institute for Czech Language CAS, Prague, http://vokabular.ujc.cas.cz.
Zikánová Š., 2008, *Postavení slovesného přísudku ve starší češtině (1500–1620)*, Prague.

Korpus tekstów staroczeskich do zastosowań lingwistycznych

S t r e s z c z e n i e

W pierwszej części artykułu zostały zaprezentowane zasady tworzenia Korpusu Języka Czeskiego Okresu Humanizmu (1500–1620). Zawiera on ponad 50 tekstów (oraz streszczeń dłuższych tekstów) długości około 600 000 słów. Mimo że ten korpus nie może być reprezentatywny, jest dobrze zrównoważony – teksty zostały wybrane pod względem typu, daty stworzenia i innych cech. Są one transkrybowane, tagi są osadzone w tekście na takiej samej zasadzie, jak w diachronicznym korpusie czeskim DIAKORP. Jedyną różnicą jest to, że korpus nie będzie lematyzowany. Nieregularne formy będą zatem oznaczane i uzupełniane przez formę podstawową, co umożliwia bardziej efektywne przeszukiwanie. Określenie form podstawowych jest oparte teoretycznie na trójpoziomowym podejściu do zjawisk leksykalnych i morfologicznych. Są one klasyfikowane jako indywidualne, zbiorowe lub systemowe. W niniejszym artykule stanowisko to jest wyjaśnione za pomocą kilku przykładów na długość samogłosek w rdzeniu i końcówkach deklinacji, jak również wariantywnych form zapożyczeń.

Zadaniem drugiej części artykułu jest prezentacja wyników analizy szyku wyrazów w humanistycznej czeszczyźnie i pokazanie możliwości eksploracji korpusu w odniesieniu do zjawisk składniowych. Uwaga jest skupiona na pozycji w zdaniu tzw. swobodnych modyfikacji wyrazowych wyrażających sposób oraz nazw aktantów.