

## Corpus-Based Study of Collocations in the AAC

**Hanno Biber, Evelyn Breiteneder, Dmitrij Dobrovolskij**

AAC (Austrian Academy Corpus), Austrian Academy of Sciences

Sonnenfelsgasse 19/8

A-1010 Vienna

Austria

Hanno.Biber@oeaw.ac.at, Evelyn.Breiteneder@oeaw.ac.at,

Dmitrij.Dobrovolskij@assoc.oeaw.ac.at

### Abstract

The Austrian Academy Corpus (AAC) is a newly founded body based within the Austrian Academy of Sciences in Vienna. Its aim is to build up a complexly structured and carefully selected collection which will include textual sources from various cultural domains of the 19<sup>th</sup> and 20<sup>th</sup> centuries. Among the variety of different texts and text types which will constitute the AAC, translations of historical literary significance will also be included. Within the framework of the AAC a Russian-German parallel corpus is planned, based on Dostoevskij's novel "The Idiot", of which there are some twenty translations into German from the Russian original. One relevant application is a bilingual study of collocations. Collocations used in Russian literature of the 19<sup>th</sup> century have so far not been intensively investigated. Most collocations used by Dostoevskij differ from the corresponding collocations of current Russian with regard to either their lexical structure or their contextual usage. These differences play an important role for current reception of the text. Comparison of the various translations will open up a new field of research as regards the presentation of collocations in bilingual lexicography.

### Preliminary Remarks

The aim of this paper is to indicate new possibilities for the investigation of linguistic properties of restricted collocations. Our approach is based upon using bilingual parallel corpora. In the first part of our paper we will describe the general framework within which these studies can be conducted and how relevant results can be achieved. This framework is provided by the Austrian Academy Corpus (AAC). In this framework among many textual resources also a small number of parallel text corpora are envisaged. One of them is the "Russian-German Parallel Corpus" based around Dostoevskij's "Idiot". In the second part of the paper we will concentrate on the specific linguistic features of Russian collocations used by Dostoevskij in the text of his novel. The focus is on the question of the translatability of these collocations. Behind this issue the general question arises whether the individual translation may claim fidelity to the original in the field of restricted collocations. Finally, the specific lexicographic consequences of this corpus-based approach will also be assessed.

### The Basis of the AAC

The Austrian Academy Corpus (AAC) is a newly founded body based within the Austrian Academy of Sciences in Vienna. Its aim is to create a large electronic text corpus and to conduct research in the field of electronic text corpora. Recently this approach also has become a research objective in the fields of textual and literary studies. The advantages of well structured digital resources for the thorough study of literary texts as well as the advantages of research based upon various textual sources of historical and cultural

significance are currently under debate in various scholarly domains. There is an ever increasing demand for such well structured and reliable resources for detailed linguistic and literary research. Such highly valuable sources are the basis not only of rigorous lexicographic work and linguistic studies but also towards the general understanding of language and of texts as cultural products.

There are a number of corpora available for linguistic and lexicographic purposes only, some of which have made valuable contributions to the advance of our knowledge about linguistic phenomena. The AAC is also designed for the study of literary texts and it will provide substantial research material for linguistic and lexicographic studies as well as for literary studies. It is planned to build a complexly structured and carefully selected collection in which textual sources from various cultural domains of the 19<sup>th</sup> and 20<sup>th</sup> centuries will be included. Among the great variety of different texts and text types that constitute the AAC, a number of translations of considerable historical literary significance will also be included. Translations, however, will constitute only a small spectrum of the overall corpus holdings and will function as experimental subcorpora for specific linguistic studies, such as the study of lexical semantics in collocations.

In general, the AAC intends to digitally present a wide selection of different sources of scholarly, journalistic and political texts which were of substantial historical and cultural influence in Austria between 1848 and 1989. Among these sources are several influential and notable journals. These journals are from various cultural and literary domains, from various historical periods within the timespan mentioned and from various regions of the German-speaking countries. The first journal to be selectively made available to a wider public will be "Der Brenner", published between 1913 and 1934 and again between 1946 and 1954 in Innsbruck in Austria. The Austrian Academy Corpus will have many other sources of equal importance, of which the widely recognized satirical journal "Die Fackel", published by Karl Kraus in Vienna between 1899 and 1936, will constitute the core and starting point for future selections of texts. For this reason selected authors of historical importance as well as selected newspaper issues from Viennese, Prague and Berlin newspapers will be included. Another journal to be integrated into the AAC is the political magazine "Die Weltbühne". It was founded in Berlin in 1905 under the name of "Die Schaubühne" by Siegfried Jacobsohn. Carl von Ossietzky, the well-known pacifist and journalist, was editor-in-chief between 1927 until 1933. After 1945 "Die Weltbühne" continued to be published for a long time in the German Democratic Republic. Other famous literary journals of the first half of the 20<sup>th</sup> century were also published in Berlin, such as "Die Aktion" and "Der Sturm", and they have manifold relations to "Die Fackel". The selection criteria for these varied resources of the AAC will not only be based upon the satirical magazine "Die Fackel" and a detailed interpretation of it, but also upon other criteria, which follow empirical, historical, pragmatic and theoretical considerations. Images and manuscripts will be included in the corpus, where necessary, because graphical and typographical information is important for meaning and interpretation. This is particularly the case for complex text structures encountered in newspapers or literary journals, and which comprise a variety of functionally different text types.

The AAC working group have been engaged in the lexicographic project FACKELLEX undertaken by the Austrian Academy of Sciences since 1991. The compilation of the phraseological dictionary, *Wörterbuch der Redensarten zu der von Karl Kraus 1899 bis 1936 herausgegebenen Zeitschrift "Die Fackel"* and its computer based working procedures and corpus based research techniques have focused the lexicographers' attention towards the possibilities and needs of digital resources. Making available well-structured corpora for such specific lexicographic purposes has been recognised as a necessary task. The text-lexicographical resources gathered at the Austrian Academy of Sciences have been and will be utilized for the compilation of the text-dictionary of the "Fackel". The mentioned phraseological dictionary [Wörterbuch 1999] was the first publication of a series of text-dictionaries about this influential literary journal. It will be followed by a *Dictionary of Insults and Invectives* as used by Karl Kraus in "Die Fackel".

The issue that the complementary corpora and lexicographic resources needed for our lexicographic undertaking of FACKELLEX were not available was also a founding motive for the AAC. The German language corpora generally available are primarily corpora which have been set up following opportunistic principles, collecting what was easily available at the time of their establishment, - hence collections of the language of newspapers of the late eighties and early nineties of the 20<sup>th</sup> century-, and have not for economic reasons catered for the vast field of literature of the past, source material which is difficult and expensive to digitise. Reliable and philologically exact corpora from the late 19<sup>th</sup> and early 20<sup>th</sup> centuries of the German language are still a desideratum. The complex task of digitizing sources from this period and integrating them into an electronic corpus will be carried out by the AAC. Special emphasis will be placed on areas that have been rather neglected in humanities computing so far. Journals and newspapers pose an especially difficult task when it comes to representation in digital form.

The Austrian Academy of Sciences signed a trilateral corpus agreement with the Berlin-Brandenburg Academy of Sciences and the Swiss Academy of Humanities and Social Sciences in March last year, which will lead to a joint effort in all three German speaking countries to develop and maintain text corpora for various purposes. The corpus established in Berlin will constitute the basis of a digital dictionary of the German language of the 20<sup>th</sup> century. These and other corpora in Germany, for instance the corpus holdings at the Institut für Deutsche Sprache in Mannheim, are unlike the AAC primarily set up for linguistic purposes.

Digital resources in the form of electronic text corpora should be regarded as structures for representing complex information. The electronic text collections established at the AAC so far and its future projects will focus on electronic representations not only of literary texts, magazines, journals and newspapers but also on a carefully considered selection of texts from several other cultural and social domains. The study and detailed investigation of texts has always been crucial for our understanding not only generally of historical processes but also of detailed linguistic features. The knowledge of texts and the accessibility of textual knowledge and of texts for linguistic, lexicographic and other related research work can be based, can only be furthered by means of massive text corpora like the AAC.

One example of the variety of different texts integrated in the AAC is a small segment, which will be build up following the rationale of a parallel corpus approach. This segment or subcorpus of the AAC will allow a more detailed investigation of relevant questions of lexical semantics and translatability in an environment which does not leave out the actual background of the historical specifics of the texts whose language is to be investigated.

### **Restricted Collocations in the “Russian-German Parallel Corpus”**

The orientation of the AAC makes it possible to conduct linguistic research in fields rather untypical for the “traditional” corpus-based approach. First, a certain diachronic depth (not only recent texts, but also texts from the 19<sup>th</sup> century) enables us to consider issues of language usage in the 19<sup>th</sup> century, and to compare it to the norms of modern usage. Second, the principle that not just German original texts are to be included in the AAC makes it possible to use translations as an empirical basis. Thus the relevant specific features of lexical units under consideration can also be analysed from a bilingual perspective.

Within the framework of the AAC, a Russian-German parallel corpus is planned, based on Dostoevskij’s “Idiot” of which there are some twenty translations into German from the Russian original. With a parallel corpus approach the base text will be aligned at the sentence level with the German translations. The general aim of the parallel corpus is the comparison of the various translations of the “Idiot” in order to test different theoretical approaches to lexical semantics, such as the hypothesis that the semantic structure of a lexical unit (word, idiom or collocation) is a context-sensitive phenomenon. Similar problems have been discussed both from a purely semantic viewpoint [Алресян 1974; 1999; Fillmore 1978; 1994; Nunberg 1978; Cruse 2000; Dobrovol’skij 2001a; 2001b] and from the perspective of NLP [Pustejovsky 1991; 1993], as well as from the point of view of cognitive linguistics and cognitive psychology [Simpson 1981; Roth & Shoben 1983; Tabossi 1988]. According to our view, the translation of a lexical unit depends not only on its denotative and significative parameters, but also on factors such as the syntactic embedding of a given word, its lexical distribution, its function in the discourse structure, as well as its communicative-pragmatic usage conditions. Only with large quantities of textual data can the behaviour of an individual word be observed within a sufficient number of authentic contexts such that plausible generalizations can be made and theory both tested and developed further.

One of the relevant applications of the parallel corpus is a bilingual study of collocations. Collocations used in 19<sup>th</sup> century Russian literature have not been intensively investigated to date. Most collocations used by Dostoevskij differ from the corresponding collocations of modern Russian with regard to either their lexical structure or their contextual usage. These differences play an important role for today’s reception of the text. The comparison of various translations will enable us to come up with innovative results regarding the presentation of collocations in bilingual lexicography.

Consider the following example.

- (1) – *Не ври пустяков*, – строго сказала Варя, которая и с князем говорила весьма сухо и только что разве вежливо. [Достоевский. *Идиот*]

The word combination *врать пустяки* (literally: “to lie nonsense”) is not possible in Russian modern usage, because now the verb *врать* has only the meaning ‘to say something that is not true, to lie’, hence it is intransitive, and cannot in principle govern any nouns. Besides the noun *пустяки* is semantically incompatible with the verb *врать*. Nevertheless, the meaning of the whole expression is absolutely clear: in some way ‘to talk nonsense’.

In this regard, an interesting theoretical question arises. From the point of view of modern usage, we are dealing with a restricted collocation, because the choice of the collocator *врать* for the base *пустяки* is neither semantically nor syntactically motivated. The word combination *врать пустяки* can be processed only as a whole. On the other hand, it is quite possible that in the 19<sup>th</sup> century this word combination was perceived as being rather free, because the verb *врать* then had a wider range of combinatorial possibilities. This means that the very notion of restricted collocation can be correctly defined only against the background of synchronically fixed semantic and combinatorial properties of every word.

#### **Towards a Dynamic Notion of Collocation**

For the lexicographical description of collocations from the 19<sup>th</sup> century and/or of collocation use by a single author (Dostoevskij, in our case) it is extremely important to develop and theoretically back up a dynamic notion of collocation which would allow us to include cases such as (1). To take a list of modern collocations as a starting point of analysis would be an incorrect approach since most interesting and non-trivial collocations (displaying specific features characteristic for a given period of language development or for the language usage by the writer in question) would be neglected in that case.

These considerations are of interest for the theory of phraseology as a whole. Traditionally, the category of collocations (or, more precisely, of restricted collocations) was defined on the basis of semantic criteria. If the choice of collocator is arbitrary to a certain extent, and cannot be explained in terms of regular rules of semantic combinatorics, we are dealing with a restricted collocation (cf. [Баранов & Добровольский 1996b; Mel’čuk 1998]). So the choice of the verb in German collocations such as *eine Entscheidung treffen* or *einen Beschluss fassen* cannot be explained in terms of semantic properties of their bases *Entscheidung* and *Beschluss*. Compare also German *Maßnahmen treffen/ergreifen* and English *to take measures*. Why do we use the verbs *treffen* and *ergreifen*, and not, e.g., *nehmen*? Why are *measures* in English *taken*, and not *met* or *grasped*?

The innovative, more objective nature of the corpus-based approach to collocation analysis results first of all from its semantic blindness. This means that all lexical cooccurrences found in the corpora more often than expected on statistical grounds are, first, considered a kind of collocations. In the next step of analysis one has to find reasons for these statistically remarkable cooccurrences (cf. [Heid 1994; Moon 1998; Steyer 2002]). Investigating restricted collocations in historical texts also demands the implementation of other tools of analysis. Not only are statistically significant cooccurrences decisive for gaining relevant data but also non-trivial features of lexical combinatorics. That is all lexical cooccurrences that violate some rules of modern usage are relevant for this domain. Even if they were not perceived as restricted collocations at the time a given text was written, they are exemplars

of lexical combinatorics not explainable in semantic terms and therefore the subject of collocation analysis. In other words, the very notion of collocation is dynamic by its nature, and can only be defined against the background of a certain set of rules concerning lexical combinatorics. If these rules change (and that is what normally happens in the field of combinatorics, because this domain is subject to change to a much higher degree than the grammar rules), then the inventory of restricted collocations will vary in accordance with this change of rules.

### A Bilingual Perspective

The bilingual perspective is helpful not only with regard to the problems of translation studies and studies of the reception of a given literary work, but also with regard to lexicographical representation of lexical units in bilingual dictionaries with a diachronic dimension.

Let us turn to the translations of (1) into German; cf. (2) and (3). Taking the two most recent translations of "The Idiot": one by Hartmut Herboth [HH 1986], and another one by Swetlana Geier [SG 1996]. These translations have determined the reception of the novel in the German-speaking world and they are comparable with each other (because they are addressing the same generation of readers).

- (2) "Rede nicht solchen Unsinn", sagte Warja streng. Sie sprach auch mit dem Fürsten in einem sehr sachlichen, gerade noch höflich zu nennenden Ton. [HH 1986]
- (3) "Mach keine dummen Witze", sagte Warja streng, die auch mit dem Fürsten sehr trocken und gerade noch höflich gesprochen hatte. [SG 1996]

For clarity, compare this passage in English translation (4).

- (4) 'Don't tell lies', said Varvara severely; she was barely polite even to the prince, and spoke extremely coldly. [AM 1992]

Contexts (2) to (4) show first of all that restricted collocations of this type can be translated rather differently, because cross-linguistic full equivalents are very rare in this area. Literal translation is rarely possible in the realm of restricted collocations, and especially in this case since, due to the striking deviation of the Russian collocation *врать пустяки* from the norms of current usage, any attempt to translate it literally word-by-word would be perceived as a violation of semantic and syntactic rules. Even the most recent translation by Swetlana Geier [SG 1996], which is known for its special emphasis on faithfulness to the original and its very careful treatment of the linguistic peculiarities of the base text, does not propose any solution as to how preserve the strangeness of the Russian collocation. Behind the problems of translatability of collocations a more general question arises, namely to what extent a translation of a fictional literary text can claim linguistic adequacy and faithfulness to the original. An empirically valid investigation of these phenomena presupposes the development of new research tools. In this regard what is needed are parallel text corpora the construction of which is now in progress at the AAC.

From a purely lexicographic viewpoint, a series of dictionaries describing all Russian collocations used in literary texts from the past would be desirable. Such a lexicographical product would make it possible to investigate historical change in this field on the one hand, and on the other to discriminate between collocations typical for a single author and collocations typical for a period of time. Let us have a look at an example.

### Non-Trivial Restricted Collocations and their Cross-Linguistic Equivalents

As mentioned above, collocations are sophisticated linguistic phenomena not only from the point of view of their translation into other languages in certain texts and their cross-linguistic equivalents in the lexical system, but also from the perspective of their dynamics in a given language. As compared to the contemporary usage rules, Dostoevskij used many collocations in a different form. For example, the collocation *принять меры* 'to take measures' (literally "to receive measures") is used in "The Idiot" in the form *взять свои меры* (literally "to take his measures"), cf. (5).

- (5) Правда, генерал, по некоторым обстоятельствам, стал излишне подозрителен; но так как он был отец и супруг опытный и ловкий, то тотчас же и *взял свои меры*. [Достоевский. Идиот]

However, this usage is not specific for Dostoevskij. The analysis of other texts from the same period shows that this collocation was used in the same form by other authors, cf. (6).

- (6) а. "Вздор, чтобы на этой квартире нельзя было жить, – думал Лугин. – Моим предшественникам, видно, не суждено было в нее перебраться – это, конечно, странно! – Но я *взял свои меры*: переехал тотчас [...]." [Лермонтов. Штосс] б. [...] мне надобно знать об участии Москвы, чтобы *взять свои меры* в рассуждении города [...]. [Дурова. Записки кавалерист-девицы]

Since this collocation is specific not just for Dostoevskij, its translation by *die entsprechenden Gegenmaßnahmen treffen* or (more freely) *eine entsprechende Position beziehen* is acceptable. Compare the following contexts.

- (7) Vielleicht hatten auch gewisse Umstände den General ein wenig zu argwöhnisch gemacht, jedenfalls *bezog* er als erfahrener und gewandter Vater und Ehemann sogleich *eine entsprechende Position*. [HH 1986]  
 (8) Allerdings war der General, gewissen Umständen zufolge, übermäßig mißtrauisch geworden; aber als erfahrener und geschickter Vater und Gatte *traf* er umgehend *die entsprechenden Gegenmaßnahmen*. [SG 1996]

Similarly, the collocation *соблюдать осторожность* is used in Dostoevskij's texts in the form *наблюдать осторожность* (cf. context (9)). This also reflects the specifics of the language usage of the 19<sup>th</sup> century. The use of the verb *наблюдать* in collocations in which the verb *соблюдать* is used in current Russian can be found, for example, in Puškin's "Queen of Spades", cf. (10).

- (9) Так как и сам Тоцкий *наблюдал* покамест, по некоторым особым обстоятельствам, чрезвычайную *осторожность* в своих шагах и только еще сондировал дело, то и родители предложили дочерям на вид только еще самые отдаленные предположения. [*Достоевский. Идиот*]
- (10) У себя принимала она весь город, *наблюдая строгий этикет* [...]. [*Пушкин. Пиковая дама*]

Both *соблюдать* and *наблюдать* – as collocators – go back to the French verb *observer*, which has both of the corresponding meanings: ‘to observe’ (cf. the Russian verb *наблюдать* in its current usage) and ≈‘to keep’ in certain collocations (cf. *соблюдать* as the bearer of the Lexical Function Oper<sub>1</sub> in the sense of Mel’čuk [1998]). The German versions of the collocation *наблюдать осторожность* in (11) and (12) do not reflect the specific features of this collocation from the perspective of current Russian, but are acceptable, because the choice of this collocation in Dostoevskij’s text is not bound to any special intentions.

- (11) Da jedoch Tozki selbst, gewisser besonderer Umstände wegen, vorerst nur sehr *behutsame Schritte unternahm* und zunächst noch sondierte, äußerten sich auch die Eltern vor ihren Töchtern nur in vagen Andeutungen. [*HH 1986*]
- (12) Da Tozkij selbst auf Grund ganz besonderer Umstände außerordentlich *vorsichtig vorging* und die Verhältnisse einstweilen bloß sondierte, stellten auch die Eltern ihren Töchtern einstweilen nur die allerentferntesten Möglichkeiten in Aussicht. [*SG 1996*]

The situation is more complicated in cases when the collocation in question is used in various forms: in its modern version and in an obsolete one. In “The Idiot” the collocation *знать себе цену* (literally “to know one’s value”) is used both in this neutral form and with the collocator *понимать* ‘to understand’. Compare (13) and (14).

- (13) Этот другой человек был во фраке, имел за сорок лет и озабоченную физиономию и был специальный, кабинетный прислужник и докладчик его превосходительства, вследствие чего и *знал себе цену*. [*Достоевский. Идиот*]
- (14) Никто не мог их упрекнуть в высокомерии и заносчивости, между тем знали, что они горды и *цену себе понимают*. [*Достоевский. Идиот*]

Linguistically, the parallel use of different forms of the same collocation shows that the relevant rule of lexical combinatorics was not fixed by that time, and the speakers had more freedom in the variation of certain expressions. The difference between the forms *знать себе цену*<sup>2</sup> and *понимать себе цену* may be important for text interpretation, therefore it seems to be desirable that the translation reflects the specific choice of the collocators in question. Interestingly, the translation of the corresponding contexts by Herboth (15-16) and by Geier (17-18) points to the semantic differences between these two collocation forms, though it is not quite clear whether these translatory differences are due to the choice of the verb in the function of Oper<sub>1</sub> in the Russian collocation or to contextual properties.



- (15) Dieser zweite, ein über vierzig Jahre alter Mann im Frack mit geschäftiger Miene, war der spezielle Kabinettsdiener und Besucheranmelder Seiner Exzellenz und *sich* in dieser Eigenschaft *seiner Wichtigkeit* wohl *bewußt*. [HH 1986]
- (16) Niemand konnte ihnen Hochmut und Anmaßung nachsagen, obwohl sie Selbstbewußtsein zeigten und offensichtlich *ihren Wert kannten*. [HH 1986]
- (17) Dieser zweite Diener trug einen Frack, mochte die Vierzig überschritten haben, und seine besorgte Miene ließ darauf schließen, daß er, der spezielle Kabinetts- und Meldediener Seiner Exzellenz, *sich seiner besonderen Würde bewußt war*. [SG 1996]
- (18) Niemand konnte ihnen Hochmut oder Dünkel vorwerfen, indes war bekannt, daß sie stolz und *sich ihres Wertes bewußt waren*. [SG 1996]

It has to be pointed out that our aim here is not to assess various translations, but to answer some purely linguistic questions. Which cross-linguistic equivalents of a given collocation are possible in principle? Which of them are actually used in texts we have at our disposal? On which factors depends the choice of equivalents? If a given collocation appears in the original text in different variants, is this variation significant for the choice of equivalents or not?

### Lexicographic Consequences

In the domain of collocations the historical change in the last 150 years is enormous. "The Idiot" contains hundreds of collocations that have fallen out of use. The implementation of a corpus-based approach will enable us to answer some lexicographically relevant questions and to pose further questions.

Firstly, a list of collocations with all their variants and modifications used in the past has to be compiled. It is a well-known fact that restricted collocations are an obstacle for automatic tagging. One possible way to solve this problem is to build exhaustive lists of collocations and idioms, which then could be automatically recognized while tagging texts.

Secondly, it would be desirable to find out which collocations are specific for a certain stage of language development, and which are specific for a single author. This would contribute to text-lexicography.

Finally, the implementation of parallel text corpora would enrich this particular lexicographical description with a bilingual perspective. One of the desirable results would be a series of bilingual dictionaries of collocations based not on vague correspondances between L<sub>1</sub> and L<sub>2</sub>, but on a large amount of serious empirical data, as provided by the Austrian Academy Corpus. In this sense, our project is just the first step in this direction.

### Endnotes

<sup>1</sup>Names are transliterated here according to international transliteration rules unlike the common popular transcription used mostly in German or English translations of Fedor Dostoevskij's work.

<sup>2</sup>The degree of idiomaticity is rather high in this case. So the expression *знать себе цену* and its obsolete variant *понимать себе цену* may be considered to be an idioms. For details about idiom usage in Dostoevskij's texts compare [Баранов & Добровольский 1996a].

### **Abbreviations**

[AM 1992] = Translation of "The Idiot" by Alan Myers.

[HH 1986] = Translation of "The Idiot" by Hartmut Herboth.

[SG 1996] = Translation of "The Idiot" by Svetlana Geier.

### **Sources**

Dostoevsky, Fyodor. *The Idiot*. Translated and edited by Alan Myers. Oxford/New York: Oxford University Press, 1992.

Dostojevskij, Fjodor. *Der Idiot*. Aus dem Russischen von Svetlana Geier. Zürich: Ammann Verlag & Co. Third edition, 1996 (1st edition 1996).

Dostojevski, Fjodor. *Der Idiot*. Aus dem Russischen von Hartmut Herboth. Berlin/Weimar: Aufbau-Verlag GmbH, 1994. In: Dostojevski, Fjodor. *Sämtliche Romane und Erzählungen*, Bde 1-11. (1st edition – in the 20-volume-edition – 1986).

Достоевский Ф.М. Идиот. In: Достоевский Ф.М. *Полное собрание сочинений в тридцати томах*, том 8. Ленинград: Наука, 1973.

### **References**

[Cruse 2000] Cruse, D.A., 2000. Lexical "Facets" between Monosemy and Polysemy, in: S. Beckmann et al. (eds.). *Sprachspiel und Bedeutung. Festschrift für Franz Hundsnurscher zum 65. Geburtstag*. Niemeyer, Tübingen.

[Dobrovol'skij 2001a] Dobrovol'skij, D.O., 2001. Pragmatische Konventionen aus kontrastiver Sicht, in: H. Schröder et al. (eds.). *Linguistik als Kulturwissenschaft. Festschrift für Bernd Spillner zum 60. Geburtstag*, pp. 31-41, Lang, Frankfurt am Main etc.

[Dobrovol'skij 2001b] Dobrovol'skij, D.O., 2001. Lexikalische Kombinatorik aus kontrastiver Perspektive, in: J. Meier & A. Ziegler (eds.). *Deutsche Sprache in Europa. Geschichte und Gegenwart. Festschrift für Ilpo Tapani Piirainen zum 60. Geburtstag*, pp. 267-283, Edition Praesens, Wien.

[Fillmore 1978] Fillmore, C.J., 1978. On the Organization of Semantic Information in the Lexicon, in: *Papers from the Parasession on the Lexicon*, pp. 148-173, Chicago Linguistic Society, Chicago.

[Fillmore 1994] Fillmore, C.J., 1994. Lexicography and Ethnographic Semantics, in: *EURALEX '94. Plenary lecture*, pp. 27-48, Vrije Universiteit, Amsterdam.

[Heid 1994] Heid, U., 1994. On Ways Words Work Together – Topics in Lexical Combinatorics, in: W. Martin et al. (eds.). *EURALEX 1994 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*, pp. 227-257, Vrije Universiteit, Amsterdam.

[Mel'čuk 1998] Mel'čuk, I., 1998. Collocations and Lexical Functions, in: A.P. Cowie (ed.). *Phraseology: Theory, Analysis, and Applications*, pp. 23-53, Clarendon Press, Oxford.

[Moon 1998] Moon, R., 1998. *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Clarendon Press, Oxford.

[Nunberg 1978] Nunberg, G., 1978. *The Pragmatics of Reference*. Indiana University Linguistics Club, Bloomington.

[Pustejovsky 1991] Pustejovsky, J., 1991. The Generative Lexicon, in: *Computational Linguistics*, 17/4, pp. 409-441, The MIT Press, Cambridge, Mass., US.

[Pustejovsky 1993] Pustejovsky, J., 1993. Type Coercion and Lexical Selection, in: J. Pustejovsky (ed.). *Semantics and the Lexicon*, pp. 73-94, Kluwer Academic Publishers, Dordrecht etc.

[Roth & Shoben 1983] Roth, E.M. & E.L. Shoben, 1983. The Effect of Context on the Structure of Categories, in: *Cognitive Psychology*, 15, pp. 346-378, Academic Press, San Diego.

- [Simpson 1981] Simpson, G.B., 1981. Meaning Dominance and Semantic Context in the Processing of Lexical Ambiguity, in: *Journal of Verbal Learning and Verbal Behavior*, 20, pp. 120-136, Academic Press, San Diego.
- [Steyer 2002] Steyer, K., 2002. Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse, in: U. Haß-Zumkehr et al. (eds.). *Ansichten zur deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. Narr, Tübingen.
- [Tabossi 1988] Tabossi, P., 1988. Accessing Lexical Ambiguity in Different Types of Sentential Context, in: *Journal of Memory and Language*, 27, pp. 324-340, Academic Press, San Diego.
- [Wörterbuch 1999] *Wörterbuch der Redensarten zu der von Karl Kraus von 1899-1936 herausgegebenen Zeitschrift "Die Fackel"*, 1999. Hrsg. von W. Welzig. Verlag der Österreichischen Akademie der Wissenschaften, Wien.
- [Апресян 1974] Апресян Ю.Д., 1974. *Лексическая семантика*. Наука, Москва.
- [Апресян 1974] Апресян Ю.Д., 1999. Принципы системной лексикографии и толковый словарь *Поэтика. История литературы. Лингвистика: Сб. к 70-летию Вяч. Вс. Иванова*, pp. 634-650, ОГИ, Москва.
- [Баранов & Добровольский 1996a] Баранов А.Н. & Д.О. Добровольский, 1996. Заметки об идиоматике Ф.М. Достоевского, in: *Слово Достоевского*, pp. 35-51, ИРЯ РАН, Москва.
- [Баранов & Добровольский 1996b] Баранов А.Н. & Д.О. Добровольский, 1996. Идиоматичность и идиомы, in: *Вопросы языкознания*, 5, pp. 51-64, Наука, Москва.

