Dissemination of principles put forward by the Explanatory Combinatorial Dictionary: three case studies

Tilmann Reuther

University of Klagenfurt
A-9020 Klagenfurt, Universitätsstrasse 65-67, Austria
tilmann.reuther@uni-klu.ac.at

Abstract

Dissemination of linguistic principles put forward by the Explanatory Combinatorial Dictionary (ECD) is a key issue for the dissemination of Meaning – Text Theory. The topic of this paper is the author’s experience with developers of teaching materials and special dictionaries with regard to the implementation of three ECD principles – (1) the lexicographic definition of a word’s meaning, (2) the government pattern of a given lexeme and (3) the collocations of a given lexeme – within the framework of three projects – (A) the Russian - German CALLex Project (1995-1996), (B) the Russian Verb Project (2004-2006) and (C) the Slovenian Military Terminology Project (2005 - 2006). Evident knowledge deficiencies among project developers and the lack of readiness to implement ECD principles in practice should be overcome with the help of a specially designed popular “Introduction to ECD principles” and an initiative to publish teaching materials and dictionaries.

Keywords

Explanatory Combinatorial Dictionary, teaching materials, dictionaries, language teaching, dissemination of Meaning – Text Theory in practice

1 The problem stated

Dissemination of linguistic principles put forward by the Explanatory Combinatorial Dictionary (ECD) is a key issue for the dissemination of Meaning – Text Theory (MTT). The author’s experience in the development of joint projects with people not directly involved in MTT shows that knowledge deficiencies among project developers and the lack of readiness to implement ECD principles in practice are still to be overcome. Over the last decade the author has taken part in three such joint projects: (A) the Russian - German CALLex Project (1995-1996), (B) the Russian Verb Project (2004-2006), and (C) the Slovenian Military Terminology Project (2005-2006). These projects involved three core ECD principles – (1) the lexicographic definition of a word’s meaning, (2) the government pattern of a given lexeme, and (3) the collocations of a given lexeme.
The CALLex Project (A) involved all three ECD principles and was based on substantial instructions and material by the scientific team leader Jurij Apresjan. The result – a lexical game for about 1000 Russian and German lexemes – was delivered to the funding INTAS agency in 1997 but in the following the project partners almost failed to present the material to a broader public. Now, it is necessary to update the material according to recent findings in Lexical Semantics and the Theory of Systematic Lexicography (by Ju.D. Apresjan), to reengineer the software package according to actual technical standards\(^1\) and to organize the publication of a CD and accompanying instructions for teachers and students of Russian and German.

The Russian Verb Project (B) involved the government patterns of 122 Russian verbs. Following the author’s reaction to a request to support the publication of a computer based teaching material – a basic course on aspect and tense of the Russian verb – by two Ukrainian colleagues, a component called “Dictionary entry” was added to the already existing program. In the course of the joint project the authors had to solve three problems: selection of only one meaning of the given verb in accordance with one government pattern, presentation of the government pattern in a relatively simple way, and furnishing the dictionary entries with appropriate examples (full sentences) for each verb. The CD and an accompanying booklet will be available soon and can be evaluated in practice. The joint project showed a need to agree on simplified lexicographic techniques and to create good linguistic examples from the spot.

The Slovenian Military Terminology Project (C) involves the collocations of a given word (and in the following the separation of different word meanings and the government patterns of lexemes) and the testing of a database for consistency. The topic emerged from a request to support a corpus and dictionary project of the Austrian Ministry of Defense and in connection with a scientific project of a former PhD student of Klagenfurt University and is still under construction. There is a need to convince the project partner of the core ECD principles and their possible application with the aim to bring the whole project to a higher level of scientific standards. One of the main difficulties is the lack of textbooks presenting ECD principles in German (or Slovenian).

In the following we will give basic information on the three projects and present some aspects for further work (Sections 2 – 4). A summary will be given in Section 5.

## 2 The CALLex Project

### 2.1 Characteristics of the CALLex game

Here we allow ourselves to go into some details of the Final Report (1997) delivered to the INTAS Agency.

\(^1\) A first step in this direction has apparently been undertaken by P. Diachenko, cf. Boguslavsky, Barrios Rodriguez & Diachenko 2006.
Section 5: Summary of the Final Report (1997, p. 13) contains the following characteristics of the project:

“CALLex is a prototype computer learning tool meant for the study of Russian and German vocabularies on the basis of innovative linguistic technologies, above all on the theory of lexical functions (LF) proposed by I. A. Mel’chuk and the theory of lexical decomposition proposed by Ju. D. Apresjan. (…)"

Fundamentally CALLex is composed of Explanatory Combinatorial Dictionaries (ECD’s) of Russian and German turned into a database which stores around two thousand lexical items (a thousand per language) with up to five types of lexicographic information assigned to each item: a) headword, b) analytical definition of its meaning in terms of semantic primitives and some intermediate senses, c) part of speech, d) translation into the other language, e) the set of LF’s defined for it, with the specification of the possible values of each LF. (…) CALLex makes use of 97 LF’s.

Five innovative linguistic games based on the concept of LF’s and analytical definitions are offered by CALLex: a) translate the given word into the other language, b) supply the values of LF’s for a given word, c) supply the values of LF’s for a given word on the basis of a translation equivalent in your mother tongue, d) supply the values of a given LF for all the keywords offered by the computer, e) guess the keyword suiting the definition offered by the computer. Depending on the complexity of LF’s all the games are arranged in three levels of difficulty. The computer makes use of a system of quantitatively assessing the user’s answers on the basis of the level of difficulty and the quality of the user’s answers.”

After these general characteristics let us now turn to what the project partners thought to be the future development of the CALLex project.

2.2 Aspects of further work on CALLex

Section 3.3: Application of results of the Final Report (1997, p. 6f.) enumerates several possible ways of further work on the project: 4

“The game will be further developed and eventually be commercialized.” This goal has not been achieved and still waits for completion. A description of the game is partly available and could be reused for this purpose.

2 INTAS has stopped its activity by January 2007 and there is no more internet access to archive materials. For general project data see http://cordis.europa.eu/search/index.cfm?fuseaction=proj.simplesdocumentlucene&HD_ID=1737007&CFID=617454&CFTOKEN=48835272
Also, no substantial part of the Final Report has ever been published otherwise. For an abstract see http://cordis.europa.eu/search/index.cfm?fuseaction=result.simplesdocumentlucene&HD_ID=4678142&CFID=617454&CFTOKEN=48835272

3 This database was programmed and filled in the ACCESS 2.0 format by German and Austrian partners.

“Means for continuing CALLex envisage a) the inclusion of other languages in CALLex, above all, English; b) expanding the ECD’s of CALLex to the size of standard lexical minimums; c) adding new types of lexicographic information, above all, syntactic information on government patterns; d) producing new linguistic games based partly on the new types of lexicographic information and partly on the use of some algebraic properties of LF’s; e) transforming the classification of lexica into the hierarchical thesaurical systems. Unfortunately, an application for a grant embodying this plan has been declined by INTAS on purely formal grounds.”

At least one part of this work program – the study and addition of government patterns for the German part (see above task c) and the testing and reengineering of the German database for consistency and completeness have been carried out in a small follow-up project at the University of Klagenfurt.5

A big part of the scientific material and instructions of the CALLex Project (written by Ju. D. Apresjan) could be reused for a future textbook on ECD principles for practical projects.6

---

5 The government patterns for German and the revised German lexicographic definitions and collocational information remained unpublished (cf. Reuther & Liaunigg 1997).


7 The upper frame contains a definition of what OPER2 is and presents two examples: unter Druck stehen (to be under pressure), Applaus bekommen (to receive applause). The correct solutions for Antwort are shown in the Help modus: eine Antwort erhalten, eine Antwort bekommen (to receive / get an answer).
3 The Russian Verb Project

3.1 Characteristics of the Russian Verb Project

The accompanying booklet (Jasnickij, Ju., I. Jasnickaja & T. Reuther 2007, Beiheft, p. 1f.) gives the following characteristics of the CD course (translation into German – T.R.):

“The computer course “The Russian Verb: Aspect. Tense. Government pattern” has been developed for a broad audience. The program functions on the Windows platform (minimum Windows 98) directly from the compact disc.

Level 1 presents an Introduction to the word forms of the Russian aspect-tense system (formation of aspectual pairs and verb forms, for 122 verbs) and concentrates on the basic semantic oppositions of the verbal aspect (process and duration vs. termination and result, repeated action vs. singular action in a concrete situation).

The effective usage of the program presupposes basic knowledge of the Russian grammatical case system (nouns and other nominal parts of speech) and a basic vocabulary of approximately 500 words.

The computer course has four parts (main menu): Textbook (Учебник), Work book (Sbornik упрацнений), Reference book (Справочник) and Dictionary (Словарь). The Reference book and the Dictionary are a database and can be used independently.

(...) The Dictionary contains manifold data on the meaning and usage of the verbs of Level 1: Translation equivalents into German and English and a Dictionary entry (Словарная статья) for every verb. By clicking on this field one gets, along with the aspectual pair, manifold semantic and syntactic data on the usage of the verb: actant positions (valence and government pattern) and example sentences. This part of the program can be used on its own as a database for the Russian verb.”

3.2 Aspects of further work on the Russian Verb Project

Level 1 of the Russian Verb Project will be available soon on CD. The authors will wait for responses from the audience and continue with Level 2 (approximately 650 new Russian verbs and more important meanings of the aspectual system, to appear by the end of 2008) and Level 3 (devoted to the so called verbs of motion in Russian). In the course of this work the joint authors will have to diversify the meaning structure of the verbs and mark explicitly different lexemes of one verb. Also, the phraseological use of some verbs will have to be addressed.
4 The Military Terminology Project

4.1 Characteristics of the Military Terminology Project

The Military Terminology Project is based on extensive corpora of the Slovenian and the Croatian languages. Here I will discuss the Slovenian part. The main author of the project (Pečnik 2003, 355) writes:

“Personal computers have become an essential tool for both linguistic studies and the practical application of languages. With a quantitative approach, i.e. a computer program for the analysis of texts, reliable data, such as phonological, morphological, lexical information, and other data, can be obtained on the characteristics of a specific type of text. Based on this information, the text elements relevant for the development of terminological data bases can be selected and also processed in a semi-automatic way. This method offers the advantage that clear and documented results can be obtained and processed, and that new linguistic challenges can be identified at an early stage. According to the criteria employed and the structure of the data base, dictionaries can be generated from such a data base.”
Dissemination of ECD principles: three case studies

At the current stage, the software for the corpus generation and processing is still in the initial DOS (as it was since the late 1990s), and the database is in MySQL. Processing in a semi-automatic way means that there is no algorithm for morphological analysis and parsing so that lemmatisation and identification of collocations is done manually on the basis of textual concordances. Also, the first dictionaries (Militärwörterbuch 2003, a printed book; Pečnik 2005, an electronic CD) were published mainly for the internal use of the Austrian and Slovenian army. The structure of the dictionary entries is still quite far away from what can be expected under an ECD approach. Cf. the following example from Pečnik (2005):

<table>
<thead>
<tr>
<th>akcija, -e, f., sam.;</th>
<th>a) Aktion, f.; b) Einsatz, c) Aktie,</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 cilj -e Einsatzziel; izvedba -e Einsatzdurchführung; začetek -e Einsatzbeginn;</td>
<td></td>
</tr>
<tr>
<td>2 diverzantska ~ Jagdkampfeinsatz; hitra ~ rascher Einsatz; kratkotrajna ~ Einsatz kurzer Dauer; posamična ~ Einzeleneinsatz; subverzivna ~ Subversionseinsatz; teroristična ~ terroristischer Einsatz;</td>
<td></td>
</tr>
<tr>
<td>3 izvajati napadni boj z ~mi den Angriffskampf mit Einsätzen führen; končati ~o den Einsatz beenden; organizirati ~o den Einsatz organisieren;</td>
<td></td>
</tr>
<tr>
<td>4 diverzantsko-teroristična ~ terroristischer Jagdkampeinsatz;</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Dictionary entry from the Slovenian Military Terminology Project. The noun akcija (action).

German equivalents of the Slovenian noun were taken from general Slovenian – German dictionaries (therefore we have meanings a) Aktion (action) and c) Aktie (stock, share) which are not found in the terminological part).

The sections 1 to 4 are quite mechanically defined as:

1 – noun + dependent key noun; 2 – single dependent adjective + key noun; 3 – verb + dependent key noun; 4 – more complex collocations of several types.

First, let us turn in more detail to section 3 of the above entry. It obviously contains values of the following lexical functions:

| Oper1 (napadni boj): izvajati napadni boj (z akcijmi) (to perform an attack (with actions)) |
| FinOper1(akcija): končati akcijo (to bring an action to the end) |
| Prepar (akcija): organizirati akcijo (to organize an action) |

Obviously, the implementation of LFs within the database will need a lot of reengineering.

Second, let us look at the dictionary entry of the verb izvajati (to perform). It is quite obviously the prototype OPER1 verb for many nouns:

<table>
<thead>
<tr>
<th>izvajati, -ti, nedov. preh. gl.;</th>
<th>a) ausführen, Verb; b) durchführen, c) vorführen, č) wahrnehmen, d) ausüben, ↑ izvesti</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ~ aktivnost die Aktivität ausüben; ~ boj den Kampf führen; (...) ~ čiščenje die Säuberung durchführen; (...) ~ dežurstvo den Dienst verrichten; (...) ~ dolžnost die Pflicht erfüllen; (...) ~ parado die Parade abhalten; (...) ~ pesem das Lied vorführen; ; (...) ~ pomoč die Hilfe leisten; (...) ~ povelje den Befehl ausführen; (...) ~ pripravo die Vorbereitung treffen; (...) ~ slovesnost den Festakt abhalten; (...) ~ ukrep die Maßnahme anwenden; (...) ~ usklajevanje die Koordinierung wahrnehmen; (...) ~ zaščito den Schutz anwenden; (...)</td>
<td></td>
</tr>
</tbody>
</table>
2 dosledno ~ konsequent anwenden; načrtovano ~ planmäßig durchführen; neprestano ~ unaufhörlich durchführen; prikrito ~ verdeckt durchführen; redno ~ regelmäßig durchführen;

3 ~ bližinski boj den Nahkampf führen; ~ bočni napad den Flankenangriff durchführen; ~ bočno zavarovanje die Flankensicherung wahrnehmen; (~) ~ delovno dolžnost die Arbeitspflicht erfüllen; ~ dežurno službo den Dienst vom Tag ausüben; ~ dozimetrično kontrola die (Strahlungs-)dosiskontrolle wahrnehmen;

4 ~ akte vodenja in poveljevanja die Anweisungen durchführen (alle Arten von Befehlen); ~ bližinski protioklepn boj den Panzernahkampf führen; (~) ~ popravljanje materialno-tehničnih sredstev die Geräte reparatur wahrnehmen; (~) ~ ukrep protidesantne zaščite die Schutzmaßnahme gegen Landungen anwenden (~)

5 ~ boj na začasno zasedenem območju das Gefecht im vorübergehend besetztem Gebiet führen; ~ boj po globini den Kampf in die Tiefe führen; (~) ~ izvodovanje s prisluskovanjem die Aufklärung mit Horchen durchführen; (~) ~ po bojnem povelju gemäß Gefechtsbefehl durchführen; ~ po določili gemäß Bestimmung ausführen; ~ po fazah phasenweise durchführen; ~ po načrtu gemäß dem Plan durchführen; plangemäß durchführen; ~ po potrebi bei Notwendigkeit führen; ~ po povelju laut Befehl durchführen; ~ po načelu napada gemäß dem Angriffsgrundsatz durchführen; (~) ~ premik s prevozom den Marsch durchführen; die Verlegung mit dem Fahrzeug durchführen; (~) ~ s posebnim bojnim sredstvom mit einer Sonderwaffe wahrnehmen; ~ s poveljem mit dem Befehl wahrnehmen; ~ s prikrito območja aus einem verdeckten Bereich aus wahrnehmen; (~) ~ v coni odgovornosti im Führungsbereich führen; ~ v gorskem svetu im Bergland durchführen; ~ v koloni in Marschkolonne durchführen; ~ v razširjeni gruppiertem v aufgelockerten Gruppierung wahrnehmen; (~) ~ v skladu s predpisom die Schutzmaßnahme gegen Landungen anwenden; ~ s posebnim bojnim sredstvom mit einer Sonderwaffe wahrnehmen; ~ s poveljem mit dem Befehl wahrnehmen; ~ s prikrito območja aus einem verdeckten Bereich aus wahrnehmen; (~) ~ v coni odgovornosti im Führungsbereich führen; ~ v gorskem svetu im Bergland durchführen; ~ v koloni in Marschkolonne durchführen; ~ v razširjeni gruppiertem v aufgelockerten Gruppierung wahrnehmen; (~) ~ v skladu s predpisom die Schutzmaßnahme gegen Landungen anwenden; ~ v skladu s programom die (Strahlungs-)dosiskontrolle wahrnehmen;

Figure 4: Dictionary entry from the Slovenian Military Terminology Project.
The verb izvajati (to perform)

German equivalents of the Slovenian verb were taken from general Slovenian-German dictionaries (therefore we do not have all translation equivalents mentioned in the terminological collocational part below, cf. führen, verrichten, erfüllen, abhalten

~ boj den Kampf führen; ~ dežurstvo den Dienst verrichten; ~ dolžnost die Pflicht erfüllen; ~ parado die Parade abhalten and some others.

The sections 1 to 5 are quite mechanically defined as:

1 – key verb + dependent noun in the accusative; 2 – adverb + key verb; 3 – key verb + dependent nominal group of adjective plus noun in the accusative; 4 – key verb + dependent nominal group of noun in the accusative plus noun in the genitive; 5 – key verb + dependent nominal group in (with) a prepositional construction.

A closer inspection reveals the massive presence of collocations of the OPER1 type, as we expected. But there are also word combinations that should not be mentioned here at all, cf.

~ v gorskem svetu im Bergland durchführen (to perform in alpine territory), ~ v vseh letnih časih zu jeder Jahreszeit durchführen (to perform in all seasons) and some others.

These word combinations as quoted here are free word combinations, not collocations.
Dissemination of ECD principles: three case studies

4.2 Aspects of further work on the Military Terminology Project

A lot of basic work has to be done to convince the main author to introduce the concept of LFs, and in consequence, to reengineer the data base. Nevertheless, this effort seems worthy of undertaking since the data of the underlying corpus are absolutely reliable and important for the study of Slovene (under the auspices of a changing political situation and the strong influence of English in the context of Slovenia’s NATO membership). Even more important will be the study of the Croatian corpora (under the auspices of a dynamically changing linguistic situation in the context of the separation of Croatia from a formerly Serbian dominated Federal Yugoslav army and state). Also, Slovenia and Croatia are quite active in lexicography but there is no trace of ECD principles either in theoretical works, or in practice.

5 Summary

The three joint projects conducted with the author’s participation over the last decade show the need of dissemination of ECD principles in an active and planned manner. There are three obvious lines of action:

- Writing and publishing introductory textbooks and instructions on ECD principles in different languages (e.g. German, Slovene, …)
- Bringing ECD principles into the broader (educational) context with the help of teaching materials and special dictionaries.8
- Starting practical work (in the classroom, and for reengineering purposes on existing electronic data) from the level of correct (simplified) government patterns and proceeding from there on to the detailed representation of the lexical meanings (and the separation of lexemes!), on the one hand, and working on collocations (and Lexical Functions - be it, for the beginning, in a simplified non-formal and/or formal notation)9, on the other hand.

Bibliography


---

8 See, e.g. Alonso Ramos (2004) as characterized by Boguslavsky, Barrios Rodriguez & Diachenko (2006, 22): “The entries of this dictionary cover the lexical field of emotions and are accompanied by a large set of exercises to be done by the learner.” For a French dictionary of a complete new type in this sense see Mel’čuk & Polguère 2007.

9 Good examples for glosses of Lexical Functions used in combination with their formal representations can be found in Boguslavsky, Barrios Rodriguez & Diachenko (2006) on the Spanish CALLEX-ESP Project, e.g., along with the notorious ‘grande’ (‘great’) for Magn, also ‘no auténtico’ (‘not authentic’) for AntiVer, ‘existir’ (‘exist’) for Func0, ‘empezar a existir’ (‘to begin to exist’) for IncepFunc0, etc.
Tilmann Reuther


