

MARIE VACHKOVÁ, PRAGUE and CYRIL BELICA, MANNHEIM

Self-Organizing Lexical Feature Maps

Semiotic Interpretation and Possible Application in Lexicography

ABSTRACT: Semiotic interpretation of lexical cohesion is a major research challenge both in theoretical and applied linguistics. From the point of view of usage-based language description, individual lexical units can be roughly characterized by their collocation profiles, i.e., by collections of condensed usage patterns extracted from very large corpora. It is posited that related lexical units tend to share portions of their respective collocation profiles. Self-organizing lexical feature maps (SOMs) have been used to analyze, cluster, and visualize the observed similarity relations among collocation profiles of written German. In this paper, the linguistic interpretation of these maps is discussed and it is shown that semiotic analysis can guide linguists in their choice of cogent indicators – those with a higher descriptive value – from a particular SOM layout. It is suggested that semiotic analysis of self-organizing lexical feature maps offers valuable insights into the collocational behavior of lexical units on the parole level. Several examples and outlined applications in practical lexicography are given.

1 Introduction

Semiotic interpretation of lexical cohesion is a major research challenge both in theoretical and applied linguistics.¹ This article² examines the semiotic value of lexical cohesion in usage-based language modeling³ and description from the point of view of practical lexicography.

We abstain from discussing the manifold theoretical approaches and their respective key concepts, both from the semiological perspective⁴ and within the framework of cognitive linguistics. For the sake of simplicity and terminological clarity, the basic Peirce's triad will be used as a reference frame throughout this article. As needed, however, we will briefly comment on

our specific approach to reconciling the contingent gap between the classical structuralist view and the theoretical status of the presented self-organizing usage-based paradigm.

This article is a work-in-progress report on an ongoing interdisciplinary discussion. The main concern here is to show how lexicographers⁵ can benefit from incorporating the semiotic analysis of self-organizing lexical feature maps (SOMs) into their standard of practice. We posit that salient SOM features stimulate the lexicographers' *associative awareness*, and, thus, encourage a guided mental imagery in an empirically founded, compelling, yet non-patronizing way.

2 Self-Organizing Lexical Feature Maps

During the past two decades, a corpus-driven profiling of various linguistic phenomena has been a prolific research concern. Based on a vast amount of data derived from very large corpora, significant progress has been achieved in the methodology of exploration and modeling of language use. For epistemic goals, however, the explanatory potential of such quantitative modeling is still poorly understood.

To study how semiotic analysis can help understand the role of such quantitative models in both descriptive and theoretical linguistics, we used the empirical baseline framework CCDB established in 2001 at the Institute for the German Language (IDS) in Mannheim. The CCDB provides a large set of empirical data derived from a 2.2 billion subset of the Mannheim GERMAN REFERENCE CORPUS (DEREKO) by performing a variety of statistical collocation analyses of the entire lemmatized corpus vocabulary. Within the CCDB, each lexical unit is characterized by its *collocation profile*, a collection of condensed usage prototypes called *higher-order collocations* and *syntagmatic patterns* extracted from the DEREKO corpus (the top lines of one such profile are shown in Figure 1). Currently, the CCDB comprises more than 220,000 nontrivial collocation profiles.

| Analysewort: schwach, Analysetyp 0 | | | |
|------------------------------------|-------|----------------------------------|--|
| ■ -1 -1 | 26409 | sozial Familien Schichten | 2 50% Familien ... sozial schwächeren Schichten |
| ■ -1 -1 | 26409 | sozial Familien Gruppen | 2 50% sozial schwache Familien ... Gruppen |
| ■ -1 -1 | 26409 | sozial Familien | 689 58% Kinder aus sozial [...] schwachen [...] Familien |
| ■ -1 -1 | 26409 | sozial Schichten Gruppen | 1 100% Gruppen ... sozial schwächsten Schichten |
| ■ -1 -1 | 26409 | sozial Schichten | 145 42% aus sozial schwachen [...] Schichten |
| ■ -1 -1 | 26409 | sozial Gruppen | 68 32% der sozial [...] schwachen Gruppen |
| ■ -1 -1 | 26409 | sozial | 2240 43% aus sozial [...] schwachen Familien ... |
| ■ 1 1 | 21350 | Trost ist allerdings | 30 76% Das ist [das] allerdings nur ein schwacher Trost |
| ■ 1 1 | 21350 | Trost ist bleibt | 5 60% bleibt ... ein schwacher Trost es ist |
| ■ 1 1 | 21350 | Trost ist | 546 84% ist [...] nur[ein] schwacher [...] Trost |
| ■ 1 1 | 21350 | Trost allerdings bleibt | 1 100% schwacher Trost bleibt ... allerdings |
| ■ 1 1 | 21350 | Trost allerdings | 48 89% Das[ist das] allerdings [nur[ein] schwacher Trost |
| ■ 1 1 | 21350 | Trost bleibt | 72 45% bleibt nur ein schwacher [...] Trost |
| ■ 1 1 | 21350 | Trost | 1773 90% nur ein schwacher [...] Trost |
| ■ 1 1 | 18082 | Konjunktur anhaltend | 46 45% der anhaltend schwachen [...] Konjunktur |
| ■ 1 1 | 18082 | Konjunktur Grund | 43 48% auf Grund ... der schwachen Konjunktur und[die ... |
| ■ 1 1 | 18082 | Konjunktur werdende | 16 93% die schwächer werdende [...] Konjunktur in ... |
| ■ 1 1 | 18082 | Konjunktur | 1966 35% der schwachen [...] Konjunktur |
| ■ 1 1 | 8786 | Leistung erschreckend Halbzeit | 3 66% Halbzeit eine erschreckend schwache Leistung |
| ■ 1 1 | 8786 | Leistung erschreckend Mannschaft | 4 75% erschreckend schwache Leistung seiner Mannschaft |
| ■ 1 1 | 8786 | Leistung erschreckend | 40 67% eine erschreckend schwache [...] Leistung |
| ■ 1 1 | 8786 | Leistung Halbzeit Mannschaft | 2 50% Mannschaft ... schwachen Leistung ... Halbzeit |
| ■ 1 1 | 8786 | Leistung Halbzeit | 30 43% der ersten Halbzeit ... eine ... ganz schwache Leistung |
| ■ 1 1 | 8786 | Leistung Mannschaft | 55 38% die schwache Leistung seiner[der] Mannschaft |
| ■ 1 1 | 8786 | Leistung | 1261 43% eine[die] schwache [...] Leistung |
| ■ 1 1 | 7099 | Glied als Kette | 34 73% als [...] schwächstes Glied in der Kette |
| ■ 1 1 | 7099 | Glied als stark | 1 100% stark ... schwächstes Glied ... als |
| ■ 1 1 | 7099 | Glied als | 103 62% die ... als [...] schwächstes [...] Glied in der |
| ■ 1 1 | 7099 | Glied Kette stark | 4 75% so stark wie das schwächste Glied in einer Kette |
| ■ 1 1 | 7099 | Glied Kette | 226 68% das schwächste Glied in der Kette |
| ■ 1 1 | 7099 | Glied stark | 57 73% nur so stark wie ihr schwächstes Glied |
| ■ 1 1 | 7099 | Glied | 586 58% das schwächste [...] Glied in der |
| ■ -1 -1 | 6806 | sehr viel | 53 54% sehr [...] viel schwächer als |
| ■ -1 -1 | 6806 | sehr Aktienbörse | 15 73% die Aktienbörse [am Mittwoch] sehr schwach geschlossen |
| ■ -1 -1 | 6806 | sehr Volumen | 23 78% Bei bei sehr schwachem Volumen hat die |
| ■ -1 -1 | 6806 | sehr | 1885 45% sehr [...] schwach |
| ■ -1 -1 | 6324 | eher vertreten | 27 88% eher [...] schwach [...] vertreten |
| ■ -1 -1 | 6324 | eher besucht | 13 100% eher schwach [...] besucht |
| ■ -1 -1 | 6324 | eher stärken | 19 50% eher schwächer als stärken |

Figure 1: Snippet of a CCDB web page showing the basic view of the collocation profile of *schwach*.

Based on the observation (cf. Keibel/Belica 2007) that related lexical units tend to share portions of their respective collocation profiles, a usage-based notion of pair-wise similarity of lexemes can be defined and quantified by measuring the degree of overlap in their corresponding profiles. For example, according to the CCDB, the twenty lexemes most strongly similar to *Quark* are *Joghurt*, *Elementarteilchen*, *Topfen*, *Proton*, *Teilchen*, *Sahne*, *Käse*, *Neutron*, *Butter*, *Sauerrahm*, *Buttermilch*, *verrühren*, *Mayonnaise*, *Frischkäse*, *Eigelb*, *Müsli*, *Atomkern*, *Milch*, *Honig* and *Milchprodukt*, respectively.

Obviously, all these lexemes are not only related to *Quark* but are also related to each other in an initially unknown, and, possibly, highly complex way. We are currently pursuing several complementary research directions that involve methodologies for discovering, visualizing,

and interpreting structures in the high-dimensional similarity space that is spanned by such inter-related collocation profiles. One of them (cf. Figure 2) uses *self-organizing feature maps* (Kohonen 1990) to arrange the lexical units in question on a two-dimensional grid such that proximity on the grid reflects the similarity between their collocation profiles⁶.

| | | | | |
|--|--|--|---|--|
| Tomate Spinat Olive marinieren Paprika Karotte Avocado püriert | Kräuter Zwiebel Meerrettich Knoblauch garnieren würfeln Fenchel gewürfelt | Pürieren Senf gehackt Essig Petersilie Parmesan Schnittlauch Olivenöl | Sauerrahm Salz g Zitronensaft vermengen dazugeben Pfeffer Obers | verrühren Esslöffel Eßlöffel unterrühren Zitronenschale vermischen aufkochen Dotter |
| Kartoffel Gurke Erdbeere Apfel Ananas Erdapfel gedünstet Blumenkohl | Sauce Kraut Zubereitung Aprikose gesalzen getrocknet Schmalz geröstet | Butter Gewürz Walnuß Walnuss Zutat Paste | Sahne gerieben Emmentaler Zimt Haselnuss Haselnuß Vanille Pflanzenöl | Eigelb Vanillezucker Schlagobers schaumig Mixer Puderzucker Staubzucker Gelatine |
| Gemüse Schinken Salat Sauerkraut gekocht Speck frischen Blutwurst | Soße Weißbrot Hülsenfrucht Ketchup Banane Reis Zitrusfrucht Dattel | Buttermilch Mayonnaise Frischkäse Honig Ei Margarine Nuss Nuß | Topfen Zucker Mehl bestreichen Teig Rosine cremig | Eiweiß |
| Salami Bratkartoffel Fleisch schmecken Rührei Pellkartoffel Leberwurst zubereitet | Käse Müsli Nudel Pudding Teigware Marmelade Vollkornbrot Schweinefleisch | Joghurt Milch Jogurt Sirup Saft Vollmilch Getreide Likör | | |
| Wurst Nachtisch Toast essen Dessert Kartoffelsalat ißt lecker | fettarm Brot Obst Wurstwaren Konfitüre Süßspeise Verzehr Gebäck | Milchprodukt Kefir Cornflakes Fruchtsaft Süßwaren Süßware Milchpulver Frischmilch | | Elementarteilchen Proton Teilchen Neutron Atomkern Elektron Wechselwirkung Spin |

Figure 2: Self-organizing lexical feature map for *Quark*.

The first tentative interpretation of the lexical feature map⁷ shown in Figure 2 (we deliberately opted for a trivial one here) might suggest that there are two distinct usage aspects, or perhaps even meanings, of *Quark* in written German, one of them being somehow related to the domain of food, while the other one to the discourse in elementary-particle physics.

In the following sections, attention is paid to the linguistic interpretation of the self-organizing feature maps in more detail. It is shown how semiotic analysis can guide linguists in their choice of lexemes' cogent descriptors – those with a high determination and discrimination value – from a particular SOM layout⁸.

3 SOM interpretation

3.1 Primary indexicality

Superficially observed, the results of any self-organizing process operating on a symbolic representation of a collocational patterning can only be construed as a synoptical arrangement of *indexical type representamens* referencing several distinct yet similar *objects* within parole: ostensibly, each lexeme shown in a lexical feature map is just a *reference* to its collocation profile, i.e., to a large number of instances of several recurrent linguistic patterns scattered across the corpus in which that lexeme occurred, and which have been statistically assessed to be prototypical of that lexeme's use. If the underlying corpus contained – to a substantial extent – utterances beyond the language competence of the interpreter (e.g., an obviously problematic idiolect), then, indeed, any semiotic analysis of a given SOM would inevitably require considerable detailed inspection of the particular corpus text spans referenced by the map or even involve consultation of secondary sources.

3.2 Re-anchoring and secondary sign stipulation

However, the larger and the more stratified the corpus, the more multifarious and, at the same time, the more general (or intersubjective) the extracted representamens. Our provisional experiments suggest that, with progressing corpus coverage, the underlying signs are interpreted by a competent – but not necessarily native (cf. Section 5) – speaker with increasing adequacy. During this interpretation, the original text-based *semiotic object* referenced by the representamen is not available to the test person. Nevertheless, instead they indigenously 're-anchor' the representamen to a substitute semiotic object in their mental lexicon. Due to such a 're-anchoring', the original sign loses its proximate indexical type, and is, strictly speaking, being replaced by another sign evincing its own sign type, i.e., *iconic*, *indexical* or *symbolic*, depending on the association triggered by the presumptive original *interpretant*. The adequacy of such a substitution can be judged by the degree of congruence of the two signs as

manifested by the topological arrangement of their representamens on the map because, as discussed above, the topological distance of any two lexemes in a particular SOM layout implies a fuzzy measure of an inherent relationship of their corresponding semiotic objects.

3.3 The SOM supersign

Note that in the foregoing paragraph it is claimed that an innate relationship should be presumed between *any* two topologically close lexemes. Hence, it follows that a *cluster* of neighbor lexemes – a particular map region – can be perceived by the cognizant subject as a self-contained representamen of its own. The corresponding new sign is referred to as *SOM supersign*, cf. Sebeok et al. 1994:1026: “Supersigns are [...] sign complexes, such as ‘propositions’, ‘sentences’, or *configurations of visual elementary signs, allowing an unique interpretation as a whole* [italics M.V., C.B.]” For example, the configuration of visual elementary signs in the right bottom corner of Figure 2, i.e., the lexeme cluster «*Elementarteilchen, Proton, Teilchen, Neutron, Atomkern, Elektron, Wechselwirkung, and Spin*» refers to a unique SOM supersign pertaining to the technical meaning of the German word *Quark*.

The representamen of a SOM supersign is not a single lexeme but rather a blurred group of topologically close lexemes, i.e., a lexeme cluster. The semiotic object of an SOM supersign is the quality of the common pair-wise relationship between its constituent lexemes. Since the quality of the underlying relationship – in the sense of ‘explication’ – is available neither in the corpus nor in the self-organizing map, an attempted semiosis, i.e., rendering the supersign meaningful, requires that its *emergent* semiotic object be sought in the point of intersection of the individual semiotic objects resulting from their *interaction* within the mental lexicon due to their connotative potential.

3.4 The two interpretation steps

This section briefly describes the proposed two-step interpretation of self-organizing lexical feature maps based on our own experience and on the feedback we received from both undergraduate and graduate students. Although the interpretation results depend on individual mental lexicon structures and on the individually conceived semiotic values, they can be communicated to and validated by other persons, which strongly suggests their intersubjective value. We maintain that this is due to the recurrence of the respective signs in the discourse as the realm of socio-cultural communication.

Step One: Identifying SOM supersigns. Start in one of the SOM corners. Look at the words in a corner square (the representamen), and, following your mental associations, try to understand the underlying lexeme clustering principle (the interpretant) within that corner as compared to the clusters in the remaining corners. Do not limit your associations to the linguistic categories, and avoid aligning them with your lexicographic expectations in step one for as long as possible. Arriving at any meaningful interpretation of the organizing criterion in a corner, you have traced out your first SOM supersign. Now, try to extend that single square cluster by inspecting its neighbor squares in all directions and reflect on the gradual shift in your associations. Then, keep looking farther at the neighbors' neighbor squares in one particular direction, and imagine your associations converging to a new 'quilting point' (cf. Lacanian *point de capiton*), and, thus, the original sign morphing into a different one. Moving your focus forth and back, try to visualize the boundary where the initial sign eventually faded out, and where a notion of a new supersign entered your mind. Repeat for all corners and all directions. Try to assign each SOM square to at least one SOM supersign.

Step Two: Interpreting SOM supersigns. Evaluate the identified SOM supersigns based on your specific lexicographic criteria of relevance. Then, map the relevant supersigns onto linguistic categories implied by your theoretical framework, e.g., on the paradigmatic level.

3.5 Linguistic interpretation of SOM supersigns

The two-dimensional SOM layout renders in a simplified way the complex multidimensional similarity relations within a set of collocation profiles. By virtue of such dimension reduction, the semiosis is visually supported, just like in the case of a diagram revealing significant mutual relations of the depicted elements. As noted in the introduction, we refer to the Peircean *icons*, *indexes (symptoms)* and *symbols*, and investigate their individual and combined nature from the lexicographer's point of view to grasp their clustering structure and to devise an appropriate strategy for their description. We do not further discuss the sign specification here, since in general, it also depends on the evaluative and individual subject criteria⁹.

Typical parole indexical signs, pronouns, appear in the SOM very rarely. They are mostly suppressed in SOM maps as they are semantically faded, and, thus, from the context and discourse point of view, do not have any significant indicative or predictive function with respect to specific autosemantic words unless they are part of a multi-word expression involved in paradigmatic, syntagmatic, and idiomatic structures. Pronouns can exceptionally be inter-

preted in their syntactic function (cf. the analysis of *genau* by Schmidt in Vachková/Schmidt/Belica 2007, where the German *w*-words – *wie*, *wann*, *wo*, etc., co-occur in the role of adverbials¹⁰). Many iconic signs (like interjections) remain manifest in self-organizing maps (e.g., *igitt*); onomatopoeic verbs may occur as a considerable part of clusters, cf. the SOM to *miauen* (not shown here) referring to the most typical domestic animals as a whole. Additionally, the mixed iconic and indexical function of each lexeme depends on the degree of phonetically-phonematically conditioned motivation, as can be seen for the verbs excerpted from the SOM to *miauen*: *blöken*, *gackern*, *krähen*, *scharren*, *schnattern*, *muhen*, *wiehern*, *grunzen*, *meckern*, *zwitschern*, *quaken*, *quieken*, *krächzen*, *fiepen*, *schnarren*, *kreischen*, *zirpen*, *piepsen*, *röhren*, *summen*, *schmatzen*, *rasseln*, *schmeicheln*, *grollen*, *gurgeln*, *flüstern*, *piepen*, *pfeifen*, *bellen*, *knurren*, *jaulen*, *fauchen*, *schnurren*, *winseln*, *maunzen*, *kratzen*, and *hecheln*. Examples of dynamic and variable shares of symbolicity (iconicity/indexicality) in a single lexeme (e.g., *Jaguar* in the SOM to *Luxus*) can be easily found in many SOM maps as well. Clusters pertaining to a closed number of utterances in the corpus can serve as navigation help to discern singular discourses in which these lexemes occur.

As discussed above, human interpreters do not reflect individual lexemes, but rather a cluster as a whole: They associate and integrate the value of every lexeme within their mental lexicon with the final abstraction label of the self-organized lexical cluster. These abstraction processes are bound to subjective epistemic operations and include introspection, evaluation and decision making during the cluster structuring and ‘reshuffling’ (i.e., redistribution of lexemes within the clusters based on their semantic affinity, cf. Appendix 1).

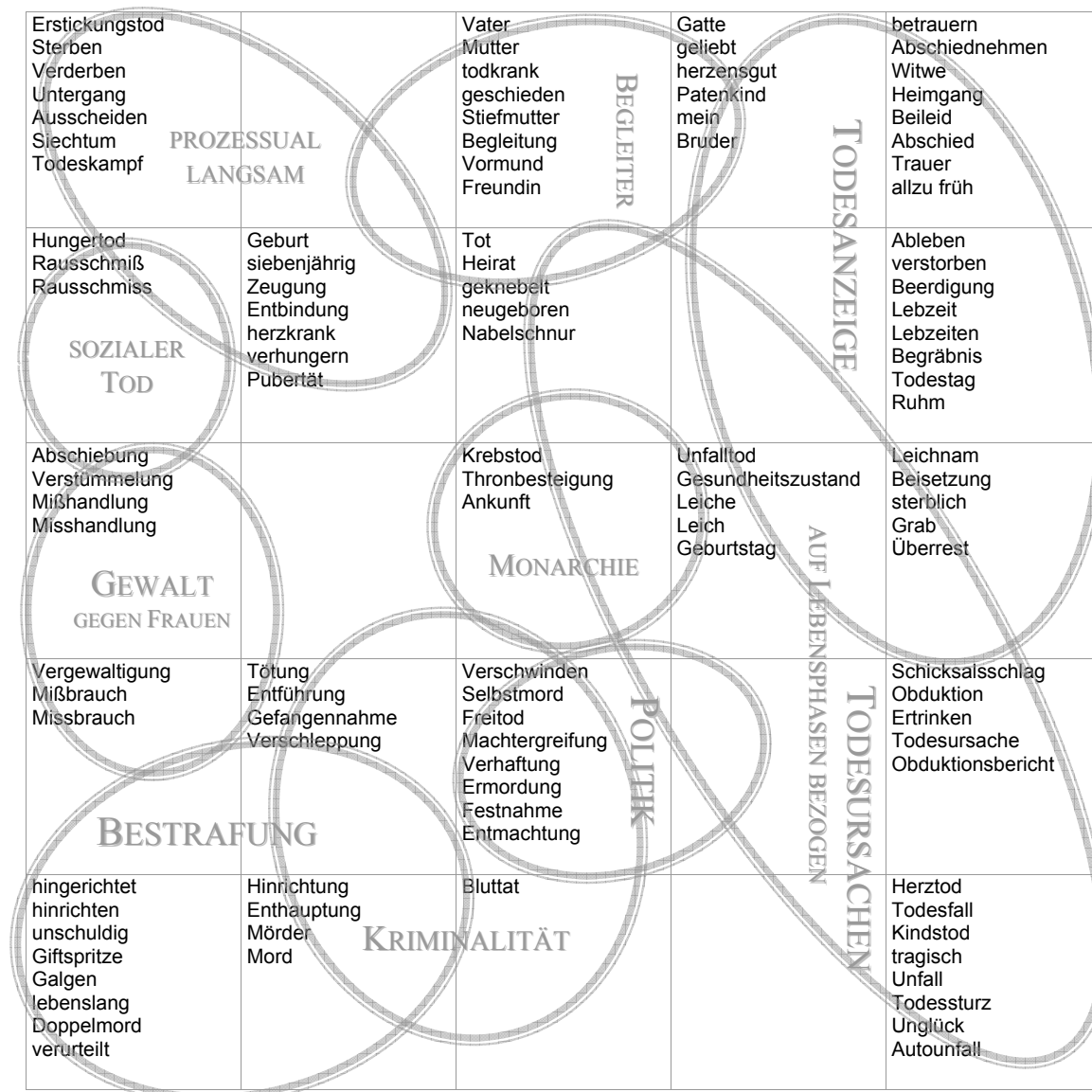


Figure 3: Annotated self-organizing lexical feature map for *Tod*.

The general organization principles perceived by the interpreter in a self-organizing map can be circumstantiated by more detailed observations involving the internal dependence paths between the individual supersign representamens as indicated by the SOM topology. (Since a considerable amount of linguistic and tacit socio-cultural knowledge is required at this stage, non-native speakers may face difficulties.) In our experience, this process reinvigorates and hones the observer's semiotic skills, and, at the same time, it stimulates the process of refinement of the initial SOM interpretation.

For example, in Figure 3, *Hungertod* in the neighborhood of *Rausschmiß/Rausschmiss* refers probably to social death. On the other hand, words like *Verstümmelung*, *Mißhandlung*, and *Misshandlung* reveal a possible reference to the women's situation in Africa. These lexemes

have reached an almost iconic character in the language of media, where the woman in Africa is often depicted as *misshandelt* or *verstümmelt*. Such statements are often stressed by expressive photographs endorsing readers' associations very intensively. Note also that the two opposite diagonal clusters in Figure 3 evoke the notions of synonymy (top left square) and hyponymy (bottom right square, with the hyperonym *Todesursache* being part of the cluster), respectively. Frequently, semiotic interpretation of SOM clusters indicates a conventionalized, formulaic expression, e.g., *Gatte, geliebt* (\rightarrow *mein geliebter Gatte* in an obituary). The lexeme *allzufrüh* also connotes the formalized utterances (cf. its significant collocates *verstorben, Gatte, Allmächtige, and Verstorbene*).

Although semiotic skills can be improved by practice, in the view of the overwhelming contextual richness of the possible semantic links, any individual differences in the eventual SOM interpretation are to be taken as a matter of fact. Arbitrary association chains resulting from inadequate, inactive, or missing links in the mental lexicon of the interpreter occur, among others, in case of polysemic words, cf. *ausscheiden: Aus dem Spiel sein, Untergang des Gewebes, Ausscheiden aus der Geschäftsführung, etc.* or *Ausscheiden* (biological function) \rightarrow *Untergang der Zellen* \rightarrow *Tod*. Understanding the key role of such metaphorical mapping in the interpretation freedom may shed some light on the entangled linking texture of the mental lexicon, and also lead to fresh lexicological views that cross the basic semantic-logical relations as cornerstone categories and theoretical frames of structural lexicology, e.g., various types of the lexical field theory (*Wortfeldtheorien*).

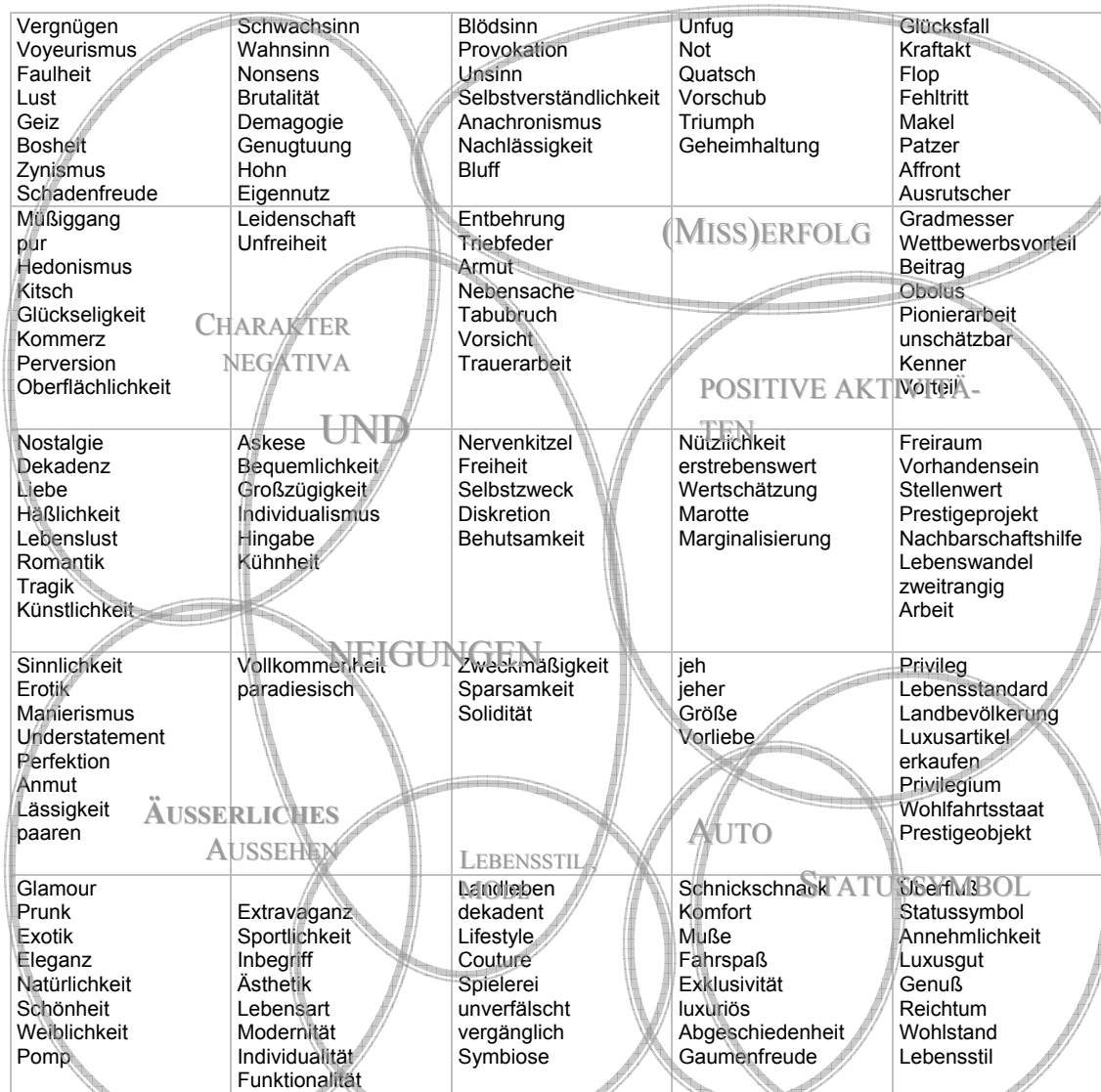


Figure 4: Annotated self-organizing lexical feature map for *Luxus*.

In Figure 4, the «CHARAKTER/NEGATIVA» cluster denotes negative moral qualities. In this environment, in the vicinity of, for example, *Faulheit* and *Lust*, the lexeme *Vergnügen* reveals the connotation ‘amusement’. Lexemes in «POSITIVE AKTIVITÄTEN» indicate a struggling and hard job leading to luxury, or at motivation (*Triebfeder* - *Armut* - *Entbehrung*). Both of these polarized clusters also hint at contextually bound facultative antonym pairs (abstract nouns *Hedonismus* - *Pionierarbeit*, *Bequemlichkeit* - *Nützlichkeit*). Neighbors in the «(MISS)ERFOLG» cluster advertize bad social behavior. The «STATUSSYMBOL» cluster characterizes *Luxus* as a status symbol related to *Überfluss* - *Reichtum* - *Annehmlichkeit* - *Wohlstand*. The icon and symbol *Auto* is adjacent to *Fahrspaß*, *Luxusgut*, to culinary relish - *Genuss*, *Gaumenfreude*, and to seclusion - *Abgeschiedenheit*. The lexemes *Schnickschnack*, *Komfort*, *Fahrspaß* indicate a luxury car. Vicinity of *Privileg* allows us to associate the social

aspect (*Landesbevölkerung - Lebensstandard - Wohlfahrtsstaat*). Lexemes clustered in the bottom part («LEBENSSTIL/MODE») associate items characteristic for luxury standards.

4 SOM-Assisted Lexicographic Analysis and Description

As mentioned above, this paper attempts to show how a traditional dictionary “*whose primary unit of description is a single word [and which thus] is singularly badly suited to explaining meaning*” (Adamska-Sałaciak 2006:47) can benefit from incorporating the results of SOM analyses. Compensating for the damage caused by “*tearing words from their mother context and setting them in rows*” (cf. Bolinger 1985:69¹¹ in Adamska-Sałaciak, *ibidem*) is the lexicographers’ main task. Usually, lexicographers interpret such statements as an invitation to create references between two or more lexemes based on standard lexicological categories in order to constitute inner relations within a dictionary text (synonymy, antonymy etc.)¹². In our opinion, this effort can be supplemented by the assistance of SOMs, as these stimulate lexicographers’ associative awareness in an unconventional way.

The semiotic analysis of self-organizing lexical feature maps can guide lexicographers in their choice of cogent linguistic phenomena and pertinent descriptive material including, but not limited to, the tailoring and evaluative revisions of the macrostructure, (e.g., compounds selection¹³), or choosing salient and compelling examples. In this process, a fundamental role is played by the inherent epistemic circularity pertaining to the autopoietic, self-organizing principle of lexical feature maps, to the emergent status of discourse-bound association potential, and to the parole-related, usage-based nature of any corpus evidence.

In the following, we focus on the abstract noun *Luxus*. We chose this word because of its particular semantic vagueness due to the differing historical, social, ethical, ethnic and confessional standards (cf. Appendix 2, Sample 1: WIKI). We show that systematic improvement in lexicographic explanations and exemplifications can be achieved by providing the entry compilers with new, SOM-based stimuli that may be unavailable to direct introspection or revealed with traditional corpus screening techniques. We use entries from the “*Duden Universalwörterbuch*” (DUW) and the “*Wahrig*” dictionaries (see samples in Appendix 2), and contrast them with the lexical material obtained from SOM analysis.

4.1 Semiotics and monolingual dictionary

Before addressing the plasticity and the semiotic value of a dictionary entry text, the traditional meaning explanation (paraphrasing) and exemplification practices in monolingual dictionaries are briefly surveyed, for once through the prism of the semiotic value of the dictionary text.

Monolingual dictionaries (*Bedeutungswörterbücher*) primarily strive to give exhaustive semantic information. However, the editors sometimes seem to virtually exhaust themselves by chiefly fabricating explanatory paraphrases and definitions according to pre-given formal principles (e.g., *luxuriös* [...]: *großen Luxus aufweisend, mit Luxus ausgestattet*). Circular definitions and tautologies are – paradoxically enough – both widely accepted and criticized, but this is not our point here. Adhering to well-formed abstract paraphrasing and other uniform formulation strategies that help explain meaning is often more straightforward than devising and using an effective strategy and specific recommendations for the choice of examples. Nevertheless, the latter seems to be burdened with the popular fallacy of being easier than the former one. Whereas formulating dictionary paraphrases is believed by many to be a highly creative, almost scientific activity, choosing a suitable example is considered a mere lexicographic skill. Both are challenging and difficult to do if done well. The semiotic value of the paraphrase is usually signaled by its prominent position ahead of examples, often in *italics*, sometimes of admirable length and level of abstractness. It is these attributes (length, style) standing for waste of time and mental pain rather than prompting for associations why less proficient dictionary users develop a tendency to assign the least priority to everything too long, complicated, or abstract (vulgo ‘*meaning nothing*’) within an entry, and to use such material only if all else failed. Consequently, it is the examples rather than the paraphrase that become the focus of their interest. To sum up, the examples should be designed as a self-contained, though rudimentary, guide for users trying to conceive meaning and the word’s usage patterns and constraints, just in the sense theoretically described above (cf. Section 3). Ideally, they would provide association stimuli targeted at each pertinent usage aspect of the word, and, at the same time, avoid any stimuli eliciting strong but misleading associations.

4.2 Paraphrases and examples in focus

In this section, we intentionally omit the discussion of the voluminous literature dealing with dictionary examples (cf., for example, Harras 1989, 1990) in order to stress a single aspect,

namely, that dictionary examples should be based on a significant connection between language usage and their discourse-bound association potential to evoke meanings effectively. In our context, we outline how the present lexicographic routine could be connected with a semiotic view to yield immediate practical consequences. In electronic dictionaries, the full text search (lemmas and full-text) makes it possible to retrieve lexemes sharing a common attribute, and to find occurrences of a word scattered throughout the whole dictionary text. This function enables the user to perceive the more or less fixed internal intertextual links. These can indicate the associations evoked and suggest a semiotic interpretation of paraphrases and examples. Selected entry texts from two dictionaries (DUW and Wahrig) have been analyzed to generate their corresponding per-dictionary token frequency lists. (We used 25 entry texts from the DUW: *Luxe, einfach, Jeunesse dorée, Komfort, Lebedame, Lebemann, Leben, lieben, luxieren, luxurieren, luxuriös, Luxus, Luxusartikel, Luxusklasse, Playgirl, Schikane, treiben, voll; delikat, luxuriös, nobel, protzig, prunkvoll, Salon*, and *treiben* and 10 entry texts from Wahrig: *Luxus, de Luxe, Fetisch, Flair, luxurieren, luxuriös, Luxusartikel, neudeutsch, Wohlleben*, and *Nobel*). The ratio of autosemantics to other morphemes (synsemantics, style markers, abbreviations, numerals, etymology, etc.) in the two lists was determined. Whereas in the DUW sample the ratio of autosemantics to other morphemes is 1.44, the Wahrig sample has the ratio of 1.71. The most frequent words (see their frequency in brackets) in the DUW sample are *Luxus* (24), *Leben* (11), *luxuriös* (10), *treiben* (6), *ausgestattet* (5), *Aufwand, großen, verschwenderisch* (4 each), *aufwendig, prunkvoll, Salon, Wagen* (3 each), the Wahrig sample contains *Luxus* (9), *Leben* (4), *Gegenstand* (3), *betreffend* (2), *deutschen* (2), *Fetisch* (2), *Flair* (2), *modischen* (2), *neudeutsche* (2), *Wohlleben* (2), *angenehm, Atmosphäre, Aufmachung, Aufwand* (each once). This simple comparison has confirmed our experience that the metalingual part in the DUW is more comprehensive in paraphrasing and in using lexicographic synonyms. Compared to Wahrig, both metalingual equipment and major differences in the example choice are striking. We put aside both strategies based on different theoretical approaches, and concentrate now on the semiotic value as a dominant quality of the entry texts to be discussed as against the semiotic quality of lexemes in SOM-clusters.

In DUW, 15 nouns (determinata) with the determinans *Luxus* are inventoried: *-artikel, -ausgabe, -auto, -dampfer, -geschöpf, -gut, -herberge, -hotel, -klasse, -limousine, -schlitten, -steuer, -villa, -weibchen, -wohnung*, from which four have a stylistic feature (*umgangssprachlich* ‘informal’ or ‘colloquial’; *scherzhaft* ‘jokingly’, *spöttisch* ‘derisively’). Four of these are names of vehicles, and four are names of housing. However, in Wahrig,

there are six nouns with the determinans *Luxus* inventoried: *Luxusartikel*, *-ausführung*, *-ausgabe*, *-ausstattung*, *-dampfer*, and *-hotel*. One of them refers to a vehicle, one to housing. The SOM offers 45 compounds: *Luxusappartement*, *-artikel*, *-ausführung*, *-ausgabe*, *-ausstattung*, *-auto*, *-bus*, *-dampfer*, *-geführt*, *-gegenstand*, *-geschöpf*, *-gut*, *-güterkonzern*, *-herberge*, *-hotel*, *-jacht*, *-karosse*, *-kategorie*, *-klasse*, *-konzern*, *-körper*, *-kreuzer*, *-leben*, *-limousine*, *-liner*, *-marke*, *-modell*, *-modernisierung*, *-problem*, *-reise*, *-restaurant*, *-sanierung*, *-schiff*, *-schlitten*, *-segment*, *-steuer*, *-suite*, *-version*, *-villa*, *-wagen*, *-wagenhersteller*, *-weibchen*, *-wohnung*, *-yacht*, and *-zug*. Five of them are names of housings, 14 denote vehicles. These results seem to harmonize with the frequent occurrence of these names in dictionary examples; *Auto*, *Wagen*, and *Sportwagen* are associated with *Luxus* quite often.

In another test, morphemes participating in the selected entries were evaluated, and differences in relations between autosemantics in the paraphrases and autosemantics in the examples were detected. The base morphemes appeared more frequently in the paraphrases than in the examples. Again, this demonstrates a strong tendency to voluminous periphrastic sentences in DUW and reinforces the typological characteristics of a monolingual dictionary, with a paraphrase or definition as the ‘*Meisterstück*’¹⁴. Out of the many observations which will be published in a separate study, we select here only a fragment pointing to the semiotic value of a paraphrase (cf. Sample 4: DUW1 in Appendix 2). Duplications and redundancies found in this entry text are of interest especially here. Redundancies can be interpreted semiotically: duplications of signs (intended or unintended) may indicate a strong internal link between *Luxus* and *verschwendend* in the mental lexicon of the entry compiler. However, the adjective which should explain the meaning is a polysemous one; its polysemy presents a typical case of a conceptual metonymy partly conventionalized and partly negotiated (in the sense of Traugott and Dasher 2005). Thus, some dictionary users may perceive from the whole entry text in the present shape a certain evaluation strain. In general, this is considered bad paraphrasing practice. Even though the noun *Luxus* bears a strong evaluative potential implicitly, this fact should not influence the paraphrase wording to such an extent (*nicht notwendig, normal*), e.g., because these lexemes in connection with *verschwendend* connote a rather negative meaning. Instead, the SOM material which embodies the associative richness of the lexical unit in its fuzziness can be consulted for more neutral advice (see examples in Sample 2: WRG1 in Appendix 2). In comparison with SOM-based suggestions, the morpheme {*verschwend*} occurs in two partial discourses: once in the ‘appearance’ cluster (as

verschwenderisch), and twice in the cluster of negative character qualities related to politics (as *Verschwendung* and *Geldverschwendung*) – see Table 1 in Appendix 1. Hierarchizing or refining connotations signaled by SOM maps becomes part of a corpus driven *modus operandi* in common usage research (i.e., common association strings). Any evaluative momentum – if present – should arise from the corpus material itself. Semiotically, a meticulous paraphrase implies fallacious 'fixedness', whereas the fuzziness of SOM highlights the dynamism of semantic change, yet moreover another observation is possible: in the same table one can detect the associations of *Luxus* which are quite usual and are demonstrated in Wahrig by the example "*sich den ~ einer Fahrkarte erster Klasse leisten*" (cf. Sample 2: WRG1 in Appendix 2). This usage, although inherent in the general definition in DUW, is captured by the SOM in a cluster in the vicinity of the lexemes *Erholung*, *Ruhepause*, *Alleinsein*, *Kurzurlaub*, *Tratsch* etc. which present the negotiated meaning of *Luxus*, namely, 'unattainable simple things in everyday stress'. The degree of abstractness of the DUW definition reveals at the same time that the lexicographer was in doubt at the instant he realized the amorphous state of lexical meaning he was expected to explicate. Sharply formulated, the compiler tries to capture an obscure and vague 'mental representation'¹⁵ which exists – if at all – not in its stable fixedness but as a 'function' of associative links built up on everyday language experience (discourse). By choosing very general examples, the compilers force the user to either follow their abstraction processes or to reflect the paraphrase, instead of presenting examples containing words like *Luxusjacht* or *Luxusrestaurant* which fit the reader's associative capacity better, and which match the colorless, though relevant collocations like *So ein Luxus; großen L. treiben; im L. leben*.

The above analysis provides the lexicographer with a stock of usage-based indicators of considerable semiotic value in compiling work or in metalexigraphic research. In this connection, lexemes in the SOMs present a promising resource for compilers of dictionaries which offer ostensive definitions. By referring to cogent indicators obtained from the SOMs, the lexicographer empowers the dictionary user to constitute meanings by activating their own association awareness. We appreciate the conclusions in Marx 1984:81 where associations are seen as a multi-tiered hierarchized phenomenon which cannot be described by pure statistics, and we encourage further investigation into SOM-based semiosis and its lexicographic applications, e.g., in the context of psycholinguistic research.

5 Intersubjectivity

Detection and interpretation of SOM clusters has been tentatively tested and discussed both in seminars and with individual undergraduate and graduate students since 2006. Our assumptions about high intersubjectivity of their interpretation have been repeatedly confirmed. Nevertheless, fine individual differences depending on individual discourse preferences, on abstraction and deduction skills, etc., remain to be explored more thoroughly. To sum up, detection of cohesive structures, ‘reshuffling’ within lexeme clusters and sign interpretation skills can be learned and developed; repeated practice leads to improved proficiency in SOM semiosis. Our observations allow generalization on several levels, mainly as follows:

- (a) Non-native German students worked out their SOM sheets with some hesitation due not so much to lack of language competence as to lack of discourse knowledge, and they needed their advisor’s guidance. Despite the mutual dependence of both these faculties, one aspect should be revisited preferably in applied linguistics and didactics (contextualization in second language acquisition, ostensive learning, creative writing, etc.).
- (b) Mapping the relevant SOM clusters (as shown, for example, in Figures 3 and 4) to classical lexicological categories describing the detected mutual lexical relations is usually quite straightforward. However, for highly heterogeneous lexical neighbors, the metalingual description of the semiotic interpretation becomes more demanding. It may entail questioning parts of the aprioristic descriptive framework and soliciting uncategorized lexical linking, and, thus, fostering creative language capabilities in various domains (e.g., poetry, verbal activity).

6 Epistemic Profit

Referring to the theory of autopoiesis, knowledge acquisition in a wide sense is more than “stor[ing] information in mental structures accessible for identification and analysis via language as a representational (denotational) sign system [...]“ (Kravchenko 2003:180)¹⁶.

Within a self organizing, autopoietic system, information “is understood as being constructed and codependent rather than instructional and referential“ (ibidem). We suggest that human language faculty is instantiated in the mind as an equilibrium between two interdependent general cognitive agencies: At the ‘fleeting initiator’ level, linguistic signs, like any other

signs, are subject to both intentional and unintentional assessments centered around the notion of similarity based on association, analogy, compulsive preference, etc., as a product of human's "interaction with the immediate environmental niche" (ibidem). By virtue of reflective classificatory and generalizing assessments, both intentional and unintentional, of the fleeting initiator level assessments of linguistic and non-linguistic signs, humans acquire, as a manifestation of innate intellectual capacity, a posteriori categorial models of their interaction with the environment, including, in particular, the structural models of the language. Along the lines of Glasersfeld's principles of radical constructivism (cf. Glasersfeld 1996:162), we regard the semiotic interpretation of self-organizing lexical feature maps as serving "... the subject's organization of the experiential world, not the discovery of an objective ontological reality".

We suggest that the proposed approach of self-organizing clustering of collocation profiles extracted from very large text corpora followed by a discourse-based semiotic interpretation and a deferred mapping of the emergent signs to systemic linguistic categories is a coherent and adequate methodology for the explorative study of language. We argue that the epistemic results obtained by applying this methodology also entail predictive and explanatory power.

7 Summary

The objective of this paper was to describe how lexicographers can benefit from incorporating semiotic analysis of self-organizing lexical feature maps into their standard of practice. Based on our experience with using SOM analysis in compiling and authoring "*The Large German-Czech Academic Dictionary*", it has been argued that salient SOM features stimulate lexicographers' associative awareness and encourage guided mental imagery leading to valuable insights into both the word semantic structure and the processes of discourse-based negotiation of lexical meaning. In seminars and with individual undergraduate and graduate students, high intersubjectivity of SOM interpretation has been observed. We encourage further interdisciplinary research into the self-organizing lexical feature maps in the semiotic context¹⁷, including follow-up studies in the domain of psycholinguistics and cognitive linguistics.

APPENDIX 1

In the CCDB (see Section 2), any self-organizing lexical feature map can be displayed in either of the two versions: ‘concise’ or ‘verbose’. The maximum total number of lexemes displayed in any SOM layout in the concise format is cut to roughly fit the screen (or page) size (cf. Figures 2, 3, and 4). In the verbose version, all pertinent lexemes are shown.

In this appendix, examples of the verbose maps for *Tod* and *Luxus* are given. For reasons of clarity, however, the synoptic view shown in Figures 3 and 4 is not presented. Instead, we only list the emergent clusters derived from the verbose map versions using semiotic analysis as discussed in this paper. In Tables 1 and 2, ‘reshuffling’ of individual lexemes pertaining to overlapping, ambiguous, or non-contiguous clusters has been carried out in addition to super-sign detection. Refined groupings marked by ‘•’ present possible ‘sub-clusters’ not only based on associations but also on synonymy, antonymy, and other detected relations. Text in square brackets is inserted to suggest possible syntactic connection between two lexemes. Style markers are placed in parentheses.

Table 1: Clusters detected in the verbose SOM of *Luxus*

Discourse: Individuelle Auffassungen des Luxus

- *Vergnügen, Spaß, Annehmlichkeit, Genuss, Wellness*
- *Glückseligkeit, Wonne, Seligkeit, Superlativ*
- *Nervenkitzel*
- *Müßiggang, Muße, Nichtstun, Erholung, Entspannung*
- *Landleben*
- *Fahrvergnügen*
- *Abgeschiedenheit, Alleinsein, Nickerchen, Ruhepause, Ruhe*
- *Kunstgenuss, Hörgenuss, Geselligkeit, Tratsch*
- *Vergnügung, Zerstreuung*
- *Gaumenfreude*
- *Wohlstand, Überfluss, Luxusgut, Ruhm*
- *Statussymbol, Garant, Freiraum, Lebensstandard, Privileg*
- *Zeitvertreib, erkaufen [durch] Stress*

- *Erinnerung [an den ersten fahrbaren] Untersatz, Triumph, Prestigeobjekt, Anblick, Lebenswandel*
- *Aufwand, Ausmaß*
- *Hype, Rummel*

Discourse: Auto

- *Lebensstil, Lifestyle, automobil*
- *Fahrvergnügen, Fahrspaß*
- *Komfort, Behaglichkeit*
- *Exklusivität, Topmodel, Einstiegsmodell, Styling*
- *Langlebigkeit, Mittelklasse, erschwinglich*
- *Nützlichkeit, Sparsamkeit, Zweckmäßigkeit, Solidität, Individualität*
- *Inbegriff, Ursprünglichkeit, Vollkommenheit, Schnelligkeit, Modernität, Ästhetik, Perfektion, Einfachheit*

Discourse: Äußerliches Aussehen (Mode, Interieur)

- *Lebensart, aristokratisch*
- *Couture, Chic, Glamour, Eleganz, Extravaganz, Schnickschnack*
- *Opulenz, Pomp, Pracht, Prunk, Glanz, Kargheit, Schlichtheit, Minimalismus, Purismus*
- ***verschwenderisch***, *dekadent, unerreicht*
- *protzen, protzig, edel, gediegen*
- *Interieur, Üppigkeit, Plüsch, Glitzer, Schnörkel, Nimbus, Protz, Glorie, orientalisches, Exotik*

Discourse: Äußerliche Wirkung, Gefühlslagen

- *Raffinesse, Noblesse*
- *Wohlklang, Schönheit, Anmut, Nüchternheit, Hauch [von] Erhabenheit*
- *Inbegriff [der] Ursprünglichkeit, Vollkommenheit, Reinheit, Modernität, Ästhetik, Originalität, Natürlichkeit, Perfektion, Einfachheit*
- *Manierismus, Raffinement, Understatement, Exzentrik, Leichtigkeit*
- *Künstlichkeit, Koketterie, Hässlichkeit*
- *Klamauk, Kitsch, Dekadenz, morbid*
- *Erotik, Sinnlichkeit*

- *Lässigkeit, Coolness, Lebensfreude, Lebenslust*
- *Spielfreude, paaren, gepaart [mit] Kühnheit, Heftigkeit, Wildheit*
- *Poesie, Pathos, Tragik, Idylle, Romantik, Abenteuerlust*
- *Nostalgie, Ausschweifung, Leidenschaft, Tristesse, pur, schier*

Discourse: Negative Neigungen, Charakterschwächen

- *Verführung, Lust, Freude, Leichtsin, Unvernunft*
- *Anarchie*
- *Pragmatismus, Opportunismus, Zynismus, Ignoranz, Egoismus, Eigennutz*
- *Voyeurismus, Sadismus, Masochismus*
- *Schadenfreude, Häme, Heuchelei, Neid, Übermut, Geldgier*
- *Bequemlichkeit, Hedonismus, Äußerlichkeit*
- *Selbstzweck, Selbstverwirklichung, Selbstverständlichkeit, Äußerlichkeit*

Discourse: Positive Neigungen

- *personifiziert[e] Liebe, Willenskraft, Freiheit, Hingabe*
- *Großzügigkeit, Enthaltensamkeit, Askese, Entbehrung, Inkarnation [der] Tüchtigkeit*
- *Weiblichkeit*

Discourse: Negativa (Ökonomie, Politik)

- *Kommerz, Kapitalismus*
- *Nonsens, Wahnsinn, Irrsinn, Hohn, Horror*
- *Dilettantismus, **Geldverschwendung**, Populismus*
- *Aktionismus, **Verschwendung**, Provokation*
- *[durch] Zufall [oder/aus] Nachlässigkeit*
- *Barbarei, Schikane, Unfug, Humbug*
- *Anachronismus, Nebensache, Tabubruch*
- *Mätzchen, Blödsinn, Schwachsinn, Unsinn, Quatsch*
- *Vorschub [leisten], Gefälligkeit*

Discourse: (Miss)erfolg

- *Flop, Glücksfall*
- *Affront, Fauxpas, Ausrutscher*
- *Patzer, Fehlgriff, Fehltritt, Fehlschuss, Makel*

- *Glücksgriff, Schildbürgerstreich*
- *Punktverlust, Ballverlust, Rückschritt*
- *Aussetzer, Reinfall, Renner, Publikumsmagnet, Knüller*
- *Bärendienst, Freundschaftsdienst*

Discourse: Positive Aktivitäten - Unternehmungen

- *Kraftakt, Eskapade*
- *Beitrag, Obolus, Kostenfaktor, Unsicherheitsfaktor*
- *Knochenarbeit, Überzeugungsarbeit, Pionierarbeit, Aufklärungsarbeit*
- *Nachbarschaftshilfe*

Table 2: Clusters detected in the verbose SOM of *Tod*

Discourse: sozialer Tod

- *Ausscheiden, Weggang, Pensionierung, Rauschmiss, Karriereende, Rauswurf*
- *künden, Abtritt*
- *ereilen, ereilt [X = etw. Unangenehmes hat jdn. ereilt]*

Discourse: Kriminalität

- *Bluttat, Ermordung*
- *Entführung, Verhaftung, Festnahme, Gefangennahme, Verschleppung, Haftentlassung, Freilassung, Tat, Attentat, Umsturz*
- *Mord, Tötung, Enthauptung, Doppelmord, Vergewaltigung, Todesschuss*
- *Mörder, Exekution, exekutieren, exekutiert, Hinrichtung, hingerichtet, hinrichten*
- *Giftspritze, Galgen*
- *Haft, Haftstrafe, Gefängnis, Zuchthaus, Todeszelle*
- *Hochverrat, Freiheitsstrafe, Zuchthausstrafe, Todesstrafe, Verbannung*
- *verurteilen, verurteilt, lebenslang, lebenslänglich*
- *Haft, Delinquent, Ehebruch, Raubmord, Vergewaltiger, Schauprozess, Rehabilitierung*

Discourse: Politik

- *Umschwung, Bestürzung, Entmachtung, Machtergreifung, Umschwung, Bestürzung, Ruhm, Scheitern*

- *Ankunft, Abreise, Heimkehr, Wintereinbruch*
- *Abgang (salopp), Ableben (gehoben), Abschied, Geheiß (gehoben), Abdankung (schweizer Deutsch: Trauerfeier)*

Discourse: Freitod

- *Verschwinden, Gesundheitszustand, Psychogramm, Selbstmord, Schicksal, Verzweiflungstat*

Discourse: KZ-Lager, Gefängnis

- *Verstümmelung, Misshandlung, Mörderin, Missbrauch, bestialisch, Mithäftling*
- *unschuldig, Gaskammer, Unschuldige, grausam*
- *Hungertod, Leid, geknebelt, verhungern, Einsamkeit*
- *schuldlos*
- *geschunden, misshandeln, misshandelt, verdursten, peinigen, gepeinigt*
- *Kopfschuss, Verhungern, Qual, Pein, Elend*

Discourse: Beerdigung

- *Beisetzung, Begräbnis, Grab, Ruhestätte, Lebzeiten, Todestag, Beileid*

Discourse: Lebensgeschichte des Verstorbenen (VIPs)

- *Thronbesteigung, Porträt, Geburtsort, Anwesenheit*
- *Erbe, Geburtshaus, Vermählung, Hochzeit*
- *Bildnis, Verehrer, Weggefährte, Geburtsstadt, Vermächtnis, Bewunderer*
- *verstorben, Heirat, Witwe, Trauer, Geliebte, Neffe, Sohn, Verlobung, Gefährtin*

Discourse: Tod in der Familie

- *Gatte, herzensgut, mein, geliebt, betrauern, betrauert, trauern, allzufrüh, Abschiednehmen, Heimgang, Anteilnahme, Hinschied*
- *Patenkind, Bruder, Cousin, Onkel, Großvater, Schwägerin, Schwiegertochter, Enkelin, Schwiegervater*
- *beweinen, Sterbebett, fürsorglich*
- *Vater, Mutter, geschieden, untreu, Stiefmutter, Freundin, außerehelich, unehelich, verheiratet, verheiraten*

Discourse: Heterogene Lebensgeschichten der Mitmenschen

- *Begleitung, Vormund, unmündig*
- *Sterben, Zeugung, neugeboren, mitansehen, Nabelschnur*
- *Pubertät, Klassenkamerad, vierjährig, Lebensende, ertränken, ertränkt*
- *Einschulung, Stiefvater, drogensüchtig*
- *Todeskampf, qualvoll, geschrien, würgen, gewürgt*
- *Liebeskummer, Siechtum, Leiden*
- *Erstickungstod, Stromstoß, Verwundung*
- *Massensterben*

Discourse: Todesursachen

- *Schicksalsschlag, tragisch*
- *Krankheit, Unfall, Unglück, Flugzeugunglück, Badeunfall, Verkehrsunfall, Autounfall, Feuertod, Todessturz, Fremdverschulden, Obduktion, Obduktionsbericht, Todesursache, Herzversagen, Herzstillstand, Herztod, Zwischenfall, Herzattack, Hirnblutung, verbluten, Erstickten, Schwächeanfall, Hirntod, Gewalteinwirkung, Gehirnblutung, Vergiftung, Unterkühlung*
- *Einlieferung, Bewusstlosigkeit, bewusstlos, erblinden, Koma, versterben*
- *Drogentod, todkrank, herzkrank, schwerkrank, aidskrank, alkoholkrank*
- *Ertrinken, Fenstersturz, schwerverletzt, verblutet*
- *Frühgeburt, Totgeburt, Säugling*

APPENDIX 2

Sample 1: WIKI

Luxus (v. lat.: *luxus* = Verschwendung, Liederlichkeit, (eigentlich) „üppige Fruchtbarkeit“) bezeichnet Verhaltensweisen, Aufwendungen oder Ausstattungen, welche über das übliche Maß (den üblichen Standard) hinausgehen bzw. über das in einer Gesellschaft als notwendig und – zum Teil auch – für sinnvoll erachtete Maß. Luxus faßt damit Phänomene zusammen, die für einen großen Teil der Bezugsgruppe zwar erstrebenswert sind, aber nicht erreichbar. Deshalb ist ihr Tauschwert oft erheblich, d.h. der Preis für ihren Erwerb hoch und deshalb sind Luxusgüter meist nur auf der Grundlage einer entsprechenden Ausstattung mit Macht oder Reichtum zu erwerben.

with the commentary

„In Ethik und Religion wird Luxus als Verschwendungssucht meist verurteilt; das Christentum betrachtet vor allem die „auffällige Zurschaustellung der Mittel“ (1. Joh 2,15) als Sünde.“

<http://de.wikipedia.org/wiki/Luxus/>

Sample 2: WRG1

>**Lu·xus** <m.; -; unz.> den normalen Lebensstandard überschreitender Aufwand, Verschwendung, Prunk; sich den ~ einer Fahrkarte erster Klasse leisten; das ist ~ das ist zwar schön u. angenehm, aber zum Leben nicht unbedingt notwendig; das ist reiner, übertriebener, totaler ~ [lat.]

Wahrig CD ROM

Sample 3: WRG2

>**lu·xu·ri'ös** <Adj.; -er, am -esten> verschwenderisch, mit großem Luxus (versehen), prunkvoll [<lat. luxuriosus “ mit frz. Endung; → Luxus]

Wahrig CD ROM

Sample 4: DUW1

Lu|xus, der; - [lat. luxus, zu lat. luxus = verrenkt u. viell. eigtl. = Abweichung vom Geraden, Normalen]: kostspieliger, **verschwenderischer**, den **normalen** Rahmen (der Lebenshaltung o. Ä.) übersteigender, nicht **notwendiger**, nur zum Vergnügen betriebener Aufwand; Pracht, **verschwenderische** Fülle: ein solches Auto ist reiner L. (ist nicht **notwendig**); großen L. treiben; im L. leben.

© Duden - Deutsches Universalwörterbuch. 4. Aufl. Mannheim 2001. [CD-ROM]

Sample 5: DUW2

lu|xu|ri|ös <Adj.> [lat. luxuriosus]: *großen Luxus aufweisend, mit Luxus ausgestattet; sehr komfortabel, prunkvoll, verschwenderisch*: eine -e Wohnung; ein -es Leben führen; der Wagen ist l. ausgestattet.

© Duden - Deutsches Universalwörterbuch. 4. Aufl. Mannheim 2001. [CD-ROM]

NOTES

- ¹ Cf. Čermák 2001:114. In our opinion, Čermák's thesis that the semiotic research in Prague does not proceed with an adequate continuity (cf. http://lexikon.meyers.de/meyers/Prager_Schule "*Durch die Verbindung von Sprach- und Literaturwissenschaft, die Arbeiten zur Stilistik, Ästhetik sowie der allgemeinen Zeichentheorie beeinflusste die Prager Schule die moderne Semiotik.*") is valid. Cf. also Vachek 2005:145 *Wortschatz – ein System?* where Havránek and Filipec (1958:177-178) encourage using empirical methods in lexicology which often misses a material base and its theorizing is considered sterile.
- ² Cf. Vachková 2007a, b and Vachková/Schmidt/Belica 2007 where the SOM approach has been described and discussed from the onomasiological and didactic views. Several related studies are currently underway.
- ³ Belica, Cyril: Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. © 2001-2007 Institut für Deutsche Sprache, Mannheim.
- ⁴ Cf. Varela/Thompson/Rosch 1991, Kravchenko 2003 and others.
- ⁵ Self-organizing lexical feature maps have been used in authoring and editing the Large German-Czech Academic Dictionary where the SOM-related evaluative discussions have been partially protocoled. However, they seem to be a promising metalexigraphic option in case of monolingual dictionaries research as well. The maps proved themselves efficient in more precise equivalency search and semantic feature description as pursuant to a typical discourse. See also Note 2 and the CNS module at <http://corpora.ids-mannheim.de/ccdb/>: "*Die laufenden Forschungsarbeiten konzentrieren sich auf die Untersuchung der Ähnlichkeit von Kookkurrenzprofilen (Modul Related Collocation Profiles), auf die Modellierung semantischer Verwandtschaft (Modul Modelling Semantic Proximity), auf die Ermittlung und Visualisierung von relevanten Gebrauchsaspekten (Modul SOM: Self-Organizing Maps) und auf die Kontrastierung von Beinahe-Synonymen (Modul Contrasting Near-Synonyms[...]*".

- ⁶ Note that, in our work, the elements subject to self-organization are the *collocation profiles* of individual lexemes. For the sake of brevity, however, we use the term ‘self-organizing lexical feature maps’ or simply ‘self-organizing maps’ (SOMs) throughout this article.
- ⁷ These are the *N* lexemes whose collocation profiles are most closely related to the collocation profile of *Quark*.
- ⁸ The specific SOM layout generally changes from run to run.
- ⁹ Cf. e.g., Keller 1995:160f.
- ¹⁰ Analyzing the difference of *genau/pünktlich* Schmidt sums up: “Bei Betrachtung der *W-Fragewörter im gelben Feld des CNS-Modells kann man problemlos schlussfolgern, welche Angaben durch Verwendung von genau präzisiert und exakter zum Ausdruck gebracht werden können und welche Satzgliedfunktion sie erfüllen (wann – Adverbiale temporis; woher – Adverbiale loci; wem – Dativobjekt usw.)*.”
- ¹¹ Bolinger, Dwight (1985): Defining the indefinable. In: Ilson, Robert I. (Ed.): *Dictionaries, lexicography and language learning*. Oxford: Pergamon Press.
- ¹² Cf. also Schippan 1992:65-71. Hermann Paul’s attempt to establish lexical network in his dictionary (1896) is an early example of dictionary texts reflecting mentally connected lexemes. However, as well as many other scholars, Paul deliberates possible associative links based on traditional paradigmatic relations and on onomasiological view.
- ¹³ For evaluating procedures concerning German compounds see Heise/Sauer 2003.
- ¹⁴ The abstract paraphrasing style in monolingual dictionaries corresponds to the German lexicographic tradition. Cf. Barz 2001, on the *idem per idem* definitions in Vachková 2007:107ff.
- ¹⁵ “[...] Augustine in his later theory of *verbum mentis* (*mental word*) is advocating the devaluation of the spoken word and the external sign in general against the internal sphere of mental cognition. It is now the mental or *interior word* (*verbum interius*), i.e., the mental concept, that is considered as word in its most proper sense, whereas the spoken word appears as a mere sign or voice of the word (*signum verbi, vox verbi*) (Augustine, *De Trinitate* XV 11 20, 1968, 486f.). Thoughts (*cogitationes*) are performed in mental words. The *verbum mentis*, corresponding to what later was called the *conceptus mentis* or *intellectus*, is by no means a ‘linguistic’ entity in the proper sense, for it is “nullius linguae”, i.e., it does not belong to any particular spoken language like Latin or Greek. So we are confronted with the paradoxical situation that linguistic terminology (e.g., *verbum, locutio, oratio, dicere*, etc.) is used to describe a phenomenon whose independence from

any language is strongly emphasized at the same time.” Stanford Encyclopedia of Philosophy at <http://plato.stanford.edu/entries/semiotics-medieval/>.

¹⁶ “*The principal activities of brains are making changes in themselves.*” (cf. Minsky, *The Society of Mind*, 1986:288 quoted in Varela/Thompson/Rosch 1991:139. For further literature cf. Varela/Thompson/Rosch 1991.

¹⁷ “[L]inguistics still remains a part of general semiotics as a science of sign systems“, cf. Frumkina 1999:31, quoted in Kravchenko 2003.

REFERENCES

Adamska-Sałaciak, Arleta. 2006. Meaning and the Bilingual Dictionary. The Case of English and Polish. *Polish Studies in English Language and Literature*. Edited by Jacek Fisiak. Vol. 18. Peter Lang.

Barz, Irmhild. 2001. Wortbildungsbeziehungen im einsprachigen Bedeutungswörterbuch. In: Korhonen, Jarmo (Hg.): *Von der mono- zur bilingualen Lexikografie für das Deutsche*. Frankfurt/ M. 2001, p. 85-100.

Čermák, František. 1994. Semiotical Lexicon and Dictionary. *Linguistica Pragensia* 1, 31-42.

Čermák, František. 1997. Semiotic Lexicon and Dictionary. *European Journal for Semiotic Studies*, Vol. 9 (3, 4).

Čermák, František. 2001. Sémiotika slovníku. In: *Lexicographica '99*, eds. S. Ondrejovič, M. Považaj, Veda Bratislava 2001, 114-125. 1. Online version at http://ucnk.ff.cuni.cz/doc/sem_slov.rtf.

Di Virgilio, Paul. 1989. Propriety Translated. Textual Affinities in the Dictionary of Semantic Universe. In: *International Journal of Lexicography*. Vol.2, No. 3. Oxford University Press, p. 259-276.

Eco, Umberto. 1991. Einführung in die Semiotik. UTB.

Glaserfeld, Ernst von. 1996. *Radical Constructivism: A Way of Knowing and Learning*. Falmer Press, London.

Harras, Gisela. 1989. Zu einer Theorie des lexikographischen Beispiels. In: Hausmann, Franz J.; Reichmann, Oskar; Wiegand, Herbert E. u. a.: *Wörterbücher, Dictionaries, Dictionnai-*

- res. Ein internationales Handbuch zur Lexikographie. p. 607-614. Berlin/New York: de Gruyter.
- Harras, Gisela. 1990. Theorie des lexikographischen Beispiels. In: Hausmann, F.-J./Wiegand, H.E./Zgusta, L. (Hrsg.): Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. p. 1003-1114 - Berlin/New York: de Gruyter.
- Harras, Gisela. 1990. Linguisten als Wörterbuchbenutzer. In: Hausmann, F.-J./Wiegand, H.E./Zgusta, L. (Hrsg.): Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie, p. 1213-1227. Berlin/New York: de Gruyter.
- Hass-Zumkehr, Ulrike. 2001. Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte. Berlin. New York. Walter der Gruyter.
- Havránek, Bohumil/Filipek, Josef. 1958. Lexikálně-sémantická výstavba hesla – ústřední otázka lexikografické práce, VPSJ, p. 177-190.
- Heise, Joachim W./Sauer, Wolfgang W. 2003. Wenn Treppenterrier zur Wortstrecke gebracht werden. Ein Vergleich von Wortschätzen in aktuellen Wörterbüchern aus dem Duden-Verlag. In: *Sprachreport* 1/2003. IDS Mannheim. p. 5-10.
- Hoffmanová, Jana. 1997. Stylistika a ... Současná situace stylistiky. Trizonia.
- Keibel, Holger/Belica, Cyril. 2007. CCDB: A Corpus-Linguistic Research and Development Workbench. In: *Proceedings of Corpus Linguistics 2007*, Birmingham. Online version at http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf.
- Keller, Rudi. 1995. Zeichentheorie. UTB 1849. Francke.
- Kohonen, Teuvo. 1990. The Self-Organizing Map. New Concepts in Computer Science: Proc. Symp. in Honour of Jean-Claude Simon, p. 181-190. Paris. AFCET.
- Kravchenko, Alexander V. 2003. The ontology of signs as linguistic and non-linguistic entities: a cognitive perspective. In: *Annual Review of Cognitive linguistics*. Vol. 1. John Benjamins. 179-191. See also <http://cogprints.org/4009/>.
- Mackenzie, Ian/Mel'čuk, Igor. 1888. Crossroads of Obsterics and Lexicography: A Case Study (The Lexicogrpahic Definition of the English Adjective *Pregnant*). In: *International Journal of Lexicography*. Vol. 1, No. 2. Oxford University Press, p. 71-83.
- Marx, Wolfgang. 1984. Das Konzept der assoziativen Bedeutung. In: Eye, Alexander v./Marx, Wolfgang (Hg.): *Semantische Dimensionen*, p. 73-84.

- Mathesius, Vilém. 1961. *Obsahový rozbor současné angličtiny*. Praha.
- O'Halloran, Kieran. 2007. *Critical Discours Analysis and the Corpus-informed Interpretation of Metaphor at the Register Level*. In: *Applied Linguistics* 28/1, p. 1-24.
- Perkuhn, Rainer/Belica, Cyril. 2006. *Korpuslinguistik – das unbekannte Wesen oder Mythen über Korpora und Korpuslinguistik*. In: *Sprachreport* 1/2006, p. 2-8. Mannheim.
- Schippa, Thea. 1992. *Lexikologie der deutschen Gegenwartssprache*. Niemeyer. Tübingen.
- Schröder, Marianne. 2002. *Brauchen wir in der Wortbildungslehre einen textlinguistischen Beschreibungsansatz, der sich an neuen Textphänomenen orientiert?* In: Fix, Ulla/Adamzik, Kirsten/Antos, Gerd/Klemm, Michael: *Brauchen wir einen neuen Textbegriff? Antworten auf eine Preisfrage*. Frankfurt am Main et al. Peter Lang.
- Sebeok, Thomas A., ed. 1994. *Encyclopedic Dicitonary of Semiotics*. Mouton De Gruyter New York.
- Šoltys, Otakar. 2004. *Komplexní komunikát a ikonizace mediální krajiny. Příloha: Jazyková komunikácia v masmediálnej sfére. (English Summary: Complex communicate and iconization of media landscape)*. In: Patráš, Vladimír, ed. 2004. *Súčasná jazyková komunikácia v interdisciplinárnych súvislostiach*. Banská Bystrica. Online version at <http://www2.fhv.umb.sk/Publikacie/2004/>
- Traugott, Elizabeth Closs/Dasher, Richard B. 2005. *Regularity in Semantic Change*. Cambridge Studies in Linguistics. Cambridge University Press.
- Vachek, Josef. 2005. *Lingvistický slovník Pražské školy. (Dictionnaire de linguistique de l'École de Prague [1960])*. Translation and edition by Vladimír Petkevič, Jaromír Tláškal. Praha. Karolinum.
- Vachková, Marie. 2007. *Kapitoly k německo-české metalexikografii I. [Kapitel zur deutsch-tschechischen Metalexikographie, Bd. I]*. Univerzita Karlova. Filozofická fakulta. ISBN 978-80-7308-199-7.
- Vachková, Marie. 2007a. *Neue Anregungen für die korpusgestützte Wortschatzforschung und die universitäre Lehre – erste Erfahrungen mit einem korpuslinguistischen Tool*. In: Berglová, Eva/Vachková, Marie/Vodrážková-Pokorná, Lenka. 2007. „...und jedes Wort hat fließende Grenzen...“. *DeskTop*. FF UK Praha. p. 41-63. In print.

- Vachková, Marie. 2007b. Korpusanalytische Methoden und Äquivalenz (am Beispiel der reihenbildenden Zweitglieder im adjektivischen Bereich). In: Heinemann, M. Liberec. TUL, p. 169-175.
- Vachková, Marie/Schmidt, Marek/Belica, Cyril. 2007. Prager Wanderungen durch die Mannheimer Quadrate. In: *Sprachreport*. März 2007. IDS Mannheim. p. 16-21.
- Vachková, Marie/Marková, Věra/Belica, Cyril. 2007. Korpusbasierte Wortschatzarbeit im Rahmen des fortgeschrittenen Germanistikunterrichts. (In print.)
- Varela, Francisco J. 1988. Cognitive Science. A Cartography of Current Ideas.
- Varela, Francisco J./Thompson, Evan/Rosch, Eleanor. 1991. The Embodied Mind. Massachusetts Institute of Technology.
- Weinreich, J. 1966. On the Semantic Structure of Language. In: Universals of Language, ed. J.H. Greenberg, Cambridge, Mass.-London, The MIT Press, p. 142-216.

BIOGRAPHICAL NOTES

Marie Vachková is Assistant Professor and Deputy Director at the Department of Germanic Studies at the Faculty of Philosophy and Arts, Charles University, Prague, Founder and Head of the Lexicographic Section, head of the project “*The Large German-Czech Academic Dictionary*“. Specialization: Lexicology and lexicography of the contemporary German language, word-formation, corpus linguistics, contrastive linguistics, written language, and translation.

Cyril Belica graduated from the Technical University in Bratislava (MSc in Computer Science). He initially worked as a research associate with the Computing Center of Slovak Academy of Sciences, before joining the Institute for the German Language in Mannheim in 1991, where he is currently Head of the Corpus Linguistics Research Program. His professional interests include lexical semantics, corpus linguistics, philosophy of linguistics, and empiricism.