Formal Concept Lattices as Semantic Maps

Daria Ryzhova and Sergei Obiedkov

National Research University Higher School of Economics,
Moscow, Russia
daria.ryzhova@mail.ru sergei.obj@gmail.com

Abstract. In this paper, we present an application for formal concept analysis (FCA) by showing how it can help construct a semantic map for a lexical typological study. We show that FCA captures typological regularities, so that concept lattices automatically built from linguistic data appear to be even more informative than traditional semantic maps. While sometimes this informativeness causes unreadability of a map, in other cases, it opens up new perspectives in the field, such as the opportunity to analyze the relationship between direct and figurative lexical meanings.

Keywords: formal concept analysis, lexical typology, semantic maps

1 Introduction

Up to now, formal concept analysis (FCA) [4] has been applied to different tasks in scientometrics [8], social network analysis [3], linguistics [5], and other fields. The application described in this paper is novel: we argue that concept lattices can serve as automatically constructed semantic maps for lexical typological studies.

We start with an explication of the task: Sec. 2 gives some background on lexical typology (namely, on the Frame approach that we adopt in this paper), and Sec. 3 introduces the notion of a semantic map. In Sec. 4, we provide basic definitions of formal concept analysis, and, in Sec. 5, we discuss possibilities of using concept lattices as semantic maps.

2 Lexical Typology

Typological studies in linguistics aim at revealing constraints and regularities in language diversity. While phonological typology deals with sets of phonemes in human languages and grammatical typology is concerned with word-forms with different grammatical functions, lexical typology analyzes how words cover the conceptual space of possible meanings.

It is well known that so-called translational equivalents have overlapping but non-identical meaning sets. For example, the English lexeme *thick* can describe both a relatively big size of an object in one of its dimensions (e.g., *thick wall*, *thick stick*) and a special consistence of substances (*thick porridge*). In Russian, two different adjectives divide this set of meanings into two parts: *tolstyj* covers the "dimensional" meaning (*tolstaja stena* 'thick wall', *tolstaja palka* 'thick stick'), and *gustoj* describes thick consistence (*gustaja kasha* 'thick porridge').

According to the Frame approach to lexical typology [6] adopted in this paper, there is a universal set of minimal meanings clustered differently in different languages. However, the number of possible clusterings is restricted, and there is a tool illustrating admissible and forbidden combinations—semantic maps.

3 Semantic Maps

A semantic map is usually represented by a connected graph with elementary lexical or grammatical meanings as nodes organized in such a way that every linguistic means (a word or an affix) covers a set of meanings inducing a connected subgraph (Semantic Map Connectivity Hypothesis, [1]). A semantic map is claimed to model a corresponding conceptual space, hence the mutual location of nodes is significant: the more often two meanings are denoted with the same linguistic means the closer they are on the map. This presupposes an additional requirement on semantic map construction that is never formulated, but is usually satisfied: there should be no edge intersections in a graph (or, in other words, the graph should be planar). Figure 1 gives a simple example, a semantic map of the lexical field 'sharp'.

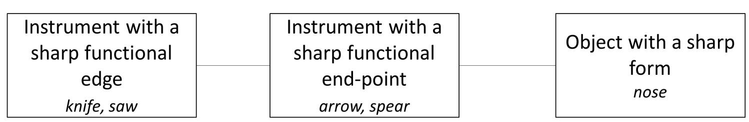

Fig. 1. Semantic map for the field 'sharp'

A semantic map suggests that any lexeme from the corresponding semantic field covers a set of contiguous meanings. According to Fig. 1, there are four possible strategies to lexicalize the field 'sharp': (1) an adjective can have all the three meanings (as the English lexeme *sharp* and the Russian *ostryj* do); (2) a word can combine the leftmost and the central meanings on the map describing thus a good functioning of an instrument (the Welsh *llym* and the Japanese *surudoi* present this pattern); (3) conversely, a lexeme can cover the central and the rightmost nodes (this is the case of the German *spitz*, French *pointu*, Kabardian *pamçe*); (4) an adjective can specialize in a single minimal meaning (compare the Aghul *hüt̄e*, the Mandarin *jianrui*, and the Welsh *pigog* describing exclusively instruments with a sharp functional edge, instruments with a sharp functional end-point, and objects with a sharp form, respectively).

It should be noted that semantic maps in lexical typology are constructed for direct meanings only, since metaphorical extensions of the words are less structured and are often considered as language-specific and escaping typological analysis.

Although semantic maps are empirically based on typological data and are not theoretically predetermined by a researcher, they are usually constructed manually. It is not difficult when dealing with a relatively small language sample and a compact semantic field (like the field 'sharp' in the example above), but it becomes much more labour-consuming with an increasing size of the dataset. However, in recent years some algorithms for automatic semantic map construction based mostly on the multidimensional scaling technique[1] have appeared [2, 9]. Here we will show that formal concept analysis is also applicable to this task and even opens up some new perspectives in the field.

4 Formal Concept Analysis

Formal concept analysis (FCA) [4] provides tools for understanding the structure of data given as a set of objects described in terms of their attributes, which is done by representing the data as a hierarchy of concepts. Every concept has extent (the set of objects that fall under the concept) and intent (the set of attributes or features that together are necessary and sufficient for an object to be an instance of the concept). Concepts are ordered in terms of being more general or less general.

We briefly introduce necessary mathematical definitions and then explain how they relate to semantic maps. Given a *(formal) context* $\mathbb{K} = (G, M, I)$, where G is called a set of *objects*, M is called a set of *attributes*, and the binary relation $I \subseteq G \times M$ specifies which objects have which attributes, the derivation operators $(\cdot)'$ are defined for $A \subseteq G$ and $B \subseteq M$ as follows:

$$A' = \{m \in M \mid \forall g \in A : gIm\}, \qquad B' = \{g \in G \mid \forall m \in B : gIm\}.$$

In words, A' is the set of attributes common to all objects of A and B' is the set of objects sharing all attributes of B.

A *(formal) concept* of the context (G, M, I) is a pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A = B'$, and $B = A'$. The set A is called the *extent* and B is called the *intent* of the concept (A, B).

A concept (A, B) is *less general* than (C, D), or is a *subconcept* of (C, D), if $A \subseteq C$. The set of all concepts ordered by this generality relation forms a lattice, called the *concept lattice* of the context $\mathbb{K}$. The concept lattice is usually visualized by a line diagram, where nodes correspond to concepts, with more general concepts placed above less general ones. Two concepts are connected by a line if one is less general than the other and there is no concept between the two. The extent of a concept can be read off by looking at the labels immediately below the corresponding node and below all nodes reachable by downward arcs.

[1] Consider, however, [7].

The intent consists of attributes indicated just above the node and those above nodes reachable by upward arcs.

In the next section, we show that by taking words as objects and frames they realize as attributes, we may interpret the concept lattice of the resulting formal context as a kind of semantic map of the corresponding lexical field.

5 Formal Concept Analysis in Lexical Typology

To construct a traditional semantic map for a lexical field, one needs a list of elementary or minimal meanings (or, in our terminology, frames[2]) and a list of words from several languages denoting these meanings. For every lexeme, a subset of frames it covers should be known. Exactly the same dataset is sufficient to build a concept lattice: the list of words is reinterpreted as a set of objects, and minimal meanings serve as attributes. Thus, typological data of this type can always be represented as a concept lattice (even though this transformation might not always result in easily interpretable structures).

A concept lattice diagram without intersections corresponds to a traditional semantic map with a linear frame configuration (a one-dimensional map). Figure 2 shows a lattice for the field 'sharp', where the mutual location of frames is exactly the same as on the traditional map (see Fig. 1). Here every node corresponds to a combination of frames (reachable via upward arcs) and a combination of words (reachable via downward arcs) realizing these frames. In terms of formal concept analysis, each node is a concept, the corresponding set of frames is its intent, and the corresponding set of words is its extent. For example, the node labeled by a set of words including "pamc̣e" corresponds to the combination of two frames, 'instrument with a sharp functional end-point' and 'object with a sharp form'; its extent contains all the words simultaneously realizing both frames (including those that realize all the three frames—these label the bottom node). A thick node corresponds to a frame combination that coincides with the entire frame set realized by some word from the semantic field (see Fig. 5 for an interesting example of a node that does not have this property).

The lattice explicitly shows the lexicalization patterns of the field: in Fig. 2, the adjectives belonging to the bottom node are dominant (they denote all the three meanings), the adjectives from the following level have two meanings each, and every lexeme from the third level specializes in a single sense. This type of data representation also highlights the most frequent lexicalization strategies: it is clear from Fig. 2 that dominant lexemes are very widespread in this semantic zone, but, if a language has two adjectives in the field 'sharp', they are more likely to draw a line between instruments with a sharp functional edge on the one side and instruments with a sharp functional end-point and objects with a sharp form on the other.

A more complex organization of a conceptual space results in edge intersections in the lattice diagram. Consider Fig. 3 representing an extended semantic

[2] Revealing the frame structure of a field is a separate task that we will not discuss here. A detailed description of this procedure is given in [6].

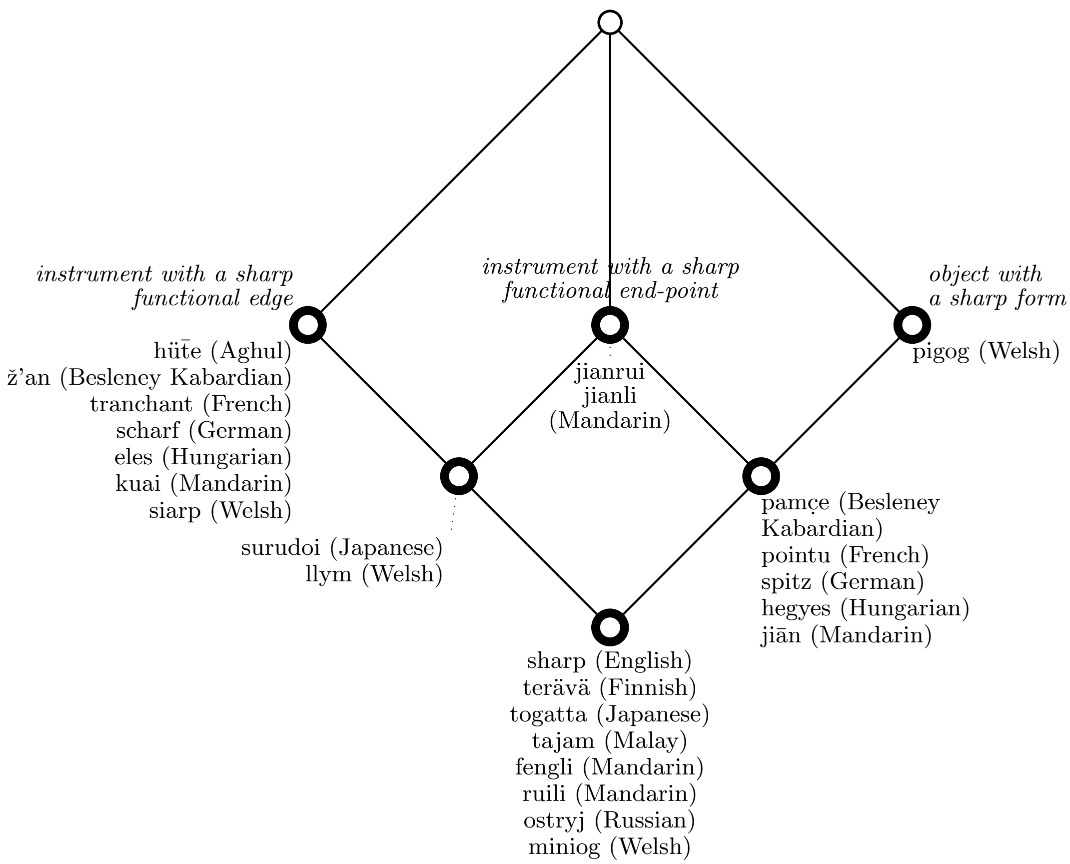

Fig. 2. Concept lattice for the field 'sharp' built from data of twelve languages

map of the field 'sharp' (with one peripheral frame included) and Fig. 4 showing the much more complicated concept lattice for the same field. Somewhat compromising on readability, the lattice diagram explicitly registers all possible frame combinations, which are implicit in the traditional semantic map.

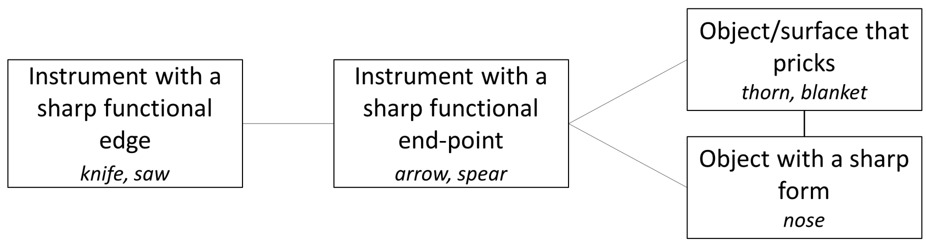

Fig. 3. Semantic map for the extended field 'sharp'

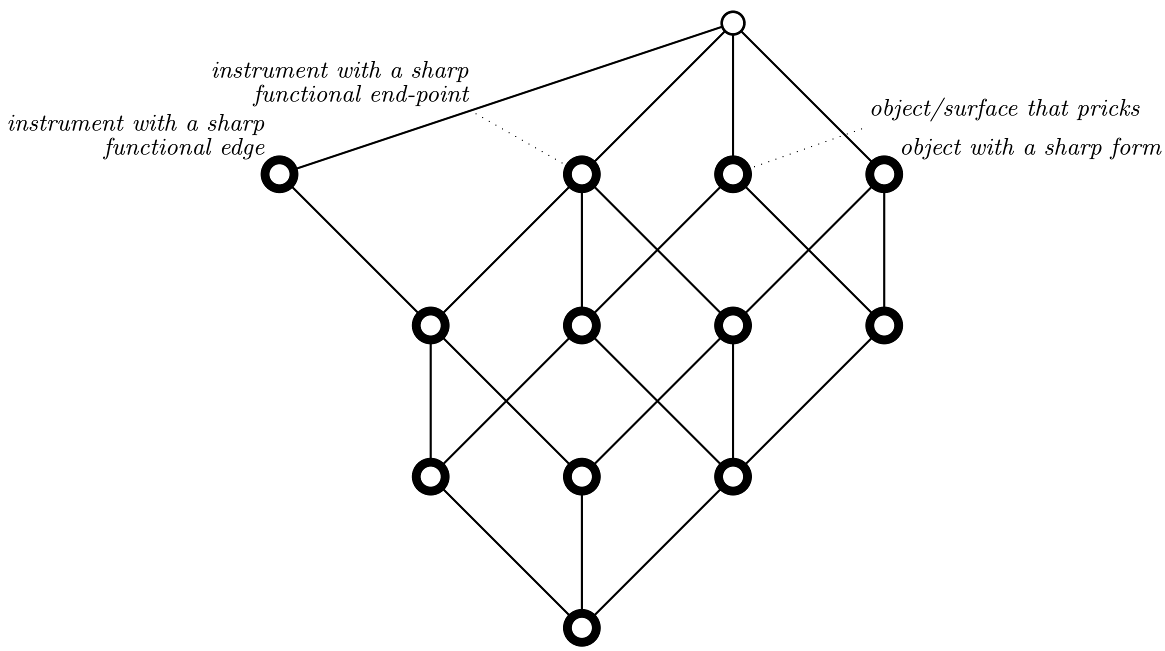

Fig. 4. Concept lattice for the extended field 'sharp' (cf. Fig. 3); words are not shown

Besides showing frame combinations, concept lattices makes explicit implicational dependencies between frames, such as "If lexeme L covers frame X, it inevitably covers frame Y, too." Traditional semantic maps consider all points of the conceptual space on par with each other. However, there are at least two situations when the lexicalization of one frame strongly depends on the realization of the other:

1. A meaning belongs to some frame, but, being far from its prototype, enables variability and can be described with the lexemes covering adjacent frames.
2. Lexicalization of a meaning is not obligatory, but, in case it is lexicalized, it always shares the lexical means with some other meaning.

Let us consider both cases in more detail.

5.1 Case 1: Transitional Microframes

According to the Frame approach to lexical typology, minimal meanings (or frames) represent different types of situations in which the lexemes in question can occur. Every frame has its prototypes, the most typical contexts. For example, the adjectives covering the frame 'instruments with a sharp functional edge' inevitably describe a sharp edge of a knife or a sharp blade, while the lexemes denoting a quality of a sharpened instrument with a functional end-point are certainly used in descriptions of sharp arrows or spears. Knife edges, blades, arrows, and spears are strongly associated with one of the minimal meanings of the field 'sharp'. However, usually there are also peripheral objects that belong to a certain frame, but, in some situations, admit another interpretation and serve as bridges between two nodes on a map. Sharp claws can serve as an example: the phrase *sharp claw* can be translated into Italian with the adjective *affilato* that is normally used with knife-type objects and with the lexeme *appuntito* that is combined with spear-type objects. This is because claws are objects with piercing end-points (and any adjective denoting the frame of piercing instruments is applicable to claws too), but they can also leave scratches and incisions, as knife-type instruments do, and in such contexts they can be described with the knife-type adjective. A concept lattice illustrating this pattern is shown in Fig. 5: the meaning 'sharp claws' is always lexicalized by words denoting 'instruments with a sharp functional end-point' and sometimes by words denoting 'instruments with a sharp functional edge', and it never possesses a special lexical means of its own.

5.2 Case 2: Metaphors

Metaphorical extensions represent the most common case of the second type of embedding. A fine-grained manual analysis of several semantic fields shows that figurative meanings are not so language-specific as they are often considered to be. On the contrary, models of semantic shifts are reproduced in different languages, regardless of genetic (un)relatedness and absence or presence of a contact between languages. For example, adjectives with the meaning 'sharp (w.r.t. cutting or piercing instruments)' describe an acute eyesight in seven languages (out of 15) of our sample: Mandarin, Japanese, Russian, Kabardian, Hungarian, Komi-Zyrian, and Malay.

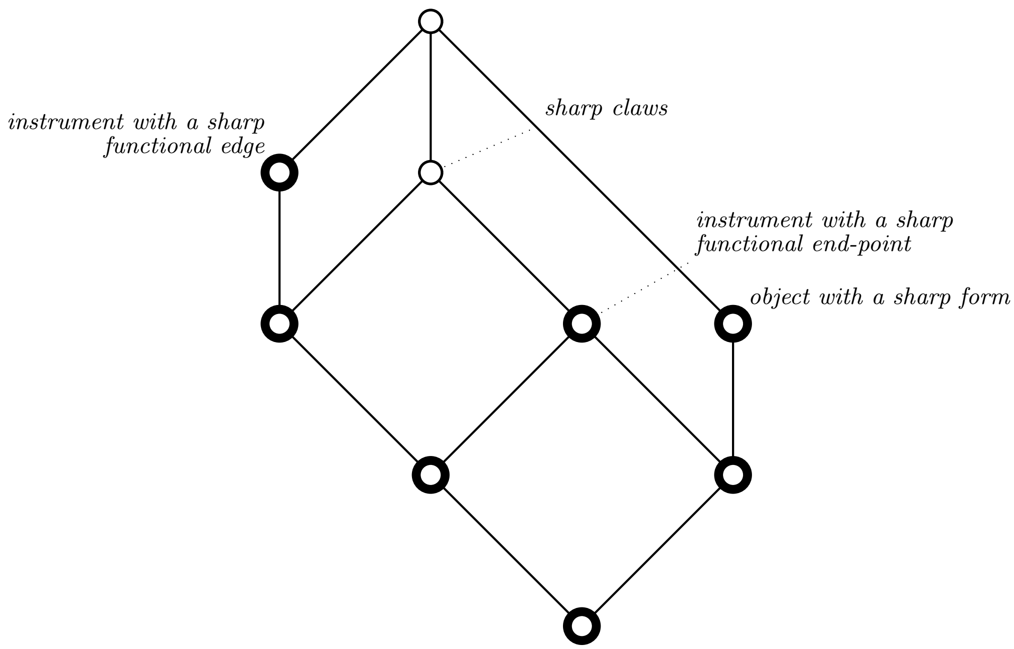

Fig. 5. Concept lattice for the field 'sharp' with a transitional microframe 'sharp claws'

Moreover, figurative meanings are usually associated with certain semantic resources, or, in our terminology, they are motivated by the direct usages. For example, a clear, precise, accurate line can be described with an adjective from the field 'sharp' (compare the English *sharp line*: *...there is a slick contrast in this drawing between the sharp black lines and the dripping green*[3]). We have found this pattern of a meaning shift in English, Japanese, Malay, and Hungarian. Some of the adjectives possessing this figurative sense are dominant (denoting all the three direct meanings), but if they cover only a subset of the direct 'sharp' frames, at least 'instrument with a sharp functional edge' is necessarily included. This association between a sharp cutting instrument and a sharp line is intuitively clear: a sharp line looks like a sharp edge.

Concept lattices successfully capture regularities of this type. Consider Fig. 6: if a lexeme has the meaning 'clear, precise, sharp (w.r.t. a line or contrast)', it inevitably describes sharp instruments with a cutting edge, too. This corresponds to the concept of 'sharp line / sharp contrast' being a subconcept of 'instrument with a sharp functional edge'; in the diagram, the node denoting the former is connected by an ascending path to the node denoting the latter.

Typology of semantic shifts remains far from being resolved. Nevertheless, formal concept analysis appears to be a powerful tool that allows verifying to what extent figurative meanings depend on the direct ones.

[3] An example is taken from a Guide for Art Students accessible at http://www.studentartguide.com/articles/line-drawings.

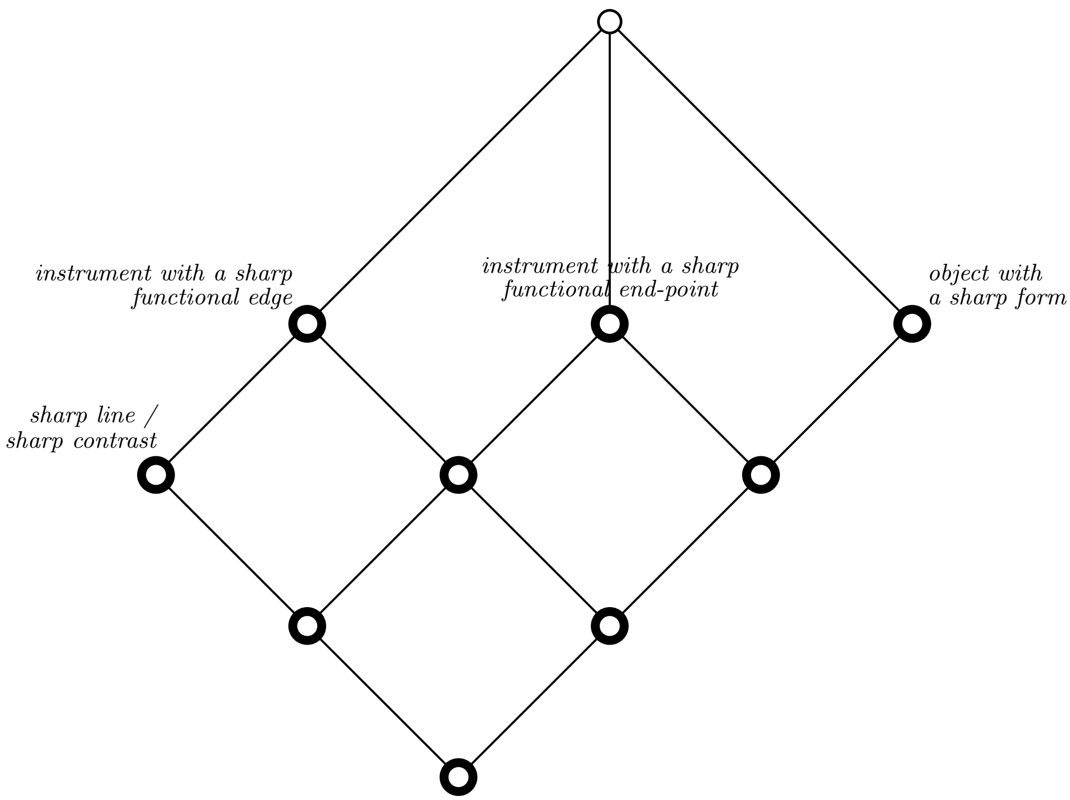

Fig. 6. Concept lattice for the field 'sharp' with selected figurative meanings

6 Conclusion

Semantic maps are usually manually constructed, and there is no universally accepted formal definition of a semantic map [2]. Concept lattices, on the other hand, provide a mathematically rigorous approach to the study of semantic fields, their structure admits unambiguous interpretation, and there are algorithms for automatic lattice construction from data on words and frames.

Unlike traditional semantic maps, concept lattices allow capturing and representing embedded structures and explicitly show frame combinations possible in the semantic field. This can sometimes make them less readable and harder to work with for a lexical typologist. In case the frame structure admits a one-dimensional representation, the correspondence between a semantic map and a concept lattice diagram is trivial (compare Figs. 1 and 2); otherwise, it is not immediately evident, because the lattice diagram records the relationship between frames only indirectly, through frame combinations in which they participate. In our future work, we hope to address this by developing representations that combine the precision and accuracy of concept lattices with readability of traditional semantic maps.

Acknowledgements

The second author is supported by the Russian Science Foundation grant 17-11-01294.

References

1. Croft, W.: Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford: Oxford University Press (2001)
2. Croft, W., Poole, K. T.: Inferring Universals from Grammatical variation: Multidimensional Scaling for Typological Analysis. Theoretical Linguistics 34.1 (2008): 1–37
3. Freeman, L.C., White D.R.: Using Galois Lattices to Represent Network Data. Sociological Methodology 23 (1993): 127–146.
4. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999)
5. Priss, U.: Linguistic Applications of Formal Concept Analysis. In Ganter, B., Stumme, G., Wille, R. (eds.): Formal Concept Analysis: Foundations and Applications. Lecture Notes in Computer Science, vol. 3626, Springer (2005): 149–160
6. Rakhilina, E., Reznikova, T.: A Frame-based Methodology for Lexical Typology. The Lexical Typology of Semantic Shifts, 58 (2016): 95–130
7. Regier, T., Khetarpal, N., Majid, A.: Inferring Semantic Maps. Linguistic Typology 17.1 (2013): 89–105
8. Roth, C., Bourgine, P.: Lattice-based Dynamic and Overlapping Taxonomies: The Case of Epistemic Communities. Scientometrics 69(2) (2006): 429–447
9. Wälchli, B., Cysouw, M.: Lexical Typology through Similarity Semantics: Toward a Semantic Map of Motion Verbs. De Gruyter 50.3 (2012): 671–710