Typological Studies in Language

The Typology of Physical Qualities

edited by Ekaterina Rakhilina,
Tatiana Reznikova and Daria Ryzhova

John Benjamins Publishing Company

The Typology of Physical Qualities

Typological Studies in Language (TSL)

A companion series to the journal Studies in Language. Volumes in this series are functionally and typologically oriented, covering specific topics in language by collecting together data from a wide variety of languages and language typologies.

For an overview of all books published in this series, please see *benjamins.com/catalog/tsl*

Editors

Volume 133

The Typology of Physical Qualities
Edited by Ekaterina Rakhilina, Tatiana Reznikova and Daria Ryzhova

The Typology of Physical Qualities

Edited by

Ekaterina Rakhilina

HSE University, Moscow | Vinogradov Russian language Institute
of the Russian Academy of Sciences

Tatiana Reznikova

HSE University, Moscow

Daria Ryzhova

HSE University, Moscow

∞™ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI z39.48-1984.

Table of contents

Introduction

The frame-based approach to the typology of qualities

Ekaterina Rakhilina[1,2] and Tatiana Reznikova[1]
[1]HSE University, Moscow / [2]Vinogradov Russian Language Institute of the
Russian Academy of Sciences

The chapter outlines the goals of our project, points out the aspects that distin-
guish the vocabulary of qualities from other lexical domains, when viewed from
a typological perspective, and introduces the methods of data collection and
analysis we use in this project and in other related studies. It goes on to discuss
the semantic parameters that motivate the lexical oppositions in various qualita-
tive domains.

Keywords: lexical typology, physical qualities, metaphor, frame, evaluation,
perception

1. Subject of this book

"What is it like?" – this is often the first question we ask when we need to in-
quire about any object; and a typical answer would be a quality word: *new, smooth,
pointed,* or *narrow.* Characteristics of things around us are a fundamental aspect
of how we conceptualize the physical world, regardless of when or where we live –
and regardless of our language. Despite this, the vocabulary of physical qualities
has received comparatively little attention in lexical typology: most of the research
so far has focused on verbs and actions they express, such as perception (Viberg
1984; Evans & Wilkins 2000; Vanhove 2008), giving (Newman (ed) 1998), motion
(Viberg 1999, 2008; Maisak & Rakhilina (eds) 2007; Driagina-Hasko & Perelmutter
(eds) 2010; Kruglyakova 2010; Wälchli & Cysouw 2012), posture (Newman (ed)
2002; Rakhilina & Lemmens 2003; Viberg 2013), cutting and breaking (Majid &
Bowerman (eds) 2007), and eating and drinking (Newman (ed) 2009). A notable
exception is color terms, cf. the classic study by Berlin & Kay 1969 which has given
rise to a full-blown school of psycholinguistic experimentation which embraces
not only color terms as such (Byrne & Hilbert 1997; Hardin & Maffi (eds) 1997;

MacLaury 1997; Gage 1999; Levinson 2000; Majid & Levinson 2007; MacLaury et al. (eds) 2007), but also other perceptual domains (see, e.g., Majid et al. 2007; Majid & Levinson (eds) 2011; Dingemanse & Majid 2012; Majid 2021). Dimensions (see Bierwisch 1967; Bierwisch & Lang 1989; Lang 2002) and temperature terms (see the recent volume Koptjevskaja-Tamm (ed) 2015) are among the other qualities that have been examined within lexico-typoligical studies.

The reason for this lack of interest in qualities probably lies in the fact that many of them heavily rely on subjective human experience. "The Princess and the Pea", the famous fairytale by Hans Christian Andersen, describes what can be called an early experiment – the princess felt uncomfortable and found her bed not soft enough because of a tiny pea placed under her twenty mattresses. In other words, the objects that may be felt as soft by some people are perceived as not so soft by others. The same is true for many other qualities; this is what makes them difficult to compare across languages if we attempt to apply the currently most widespread method of lexico-typological research, i.e. psycholinguistic experiments, which yield brilliant results in many other cases.

The basic principle of such experiments is to expose interviewees to various objects and ask them to describe the objects in their language's terms. However, if perception of most qualities is highly subjective, we need a massive sample to reach a reliable consensus, and even that does not guarantee that the results are going to be comparable across languages. Thus, perception of physical qualities appears to be too ambiguous and amorphous to be tackled with purely experimental methods.[1] It is presumably this reason that moved the Nijmegen school of lexical typology (which has until recently been the foremost purveyor of experimental techniques) to be shifting its emphasis towards polymethodological approaches (cf. Kopecka & Narasimhan (eds) 2012; Majid 2015).

We encountered a similar problem with verbs when attempting a cross-linguistic description of pain predicates – they are likewise difficult to study in a linguistic experiment. Indeed, one could hardly imagine sticking pins into speakers of different languages or pinching them, and even if such a dubious experiment did take place, its results would most probably be of little value due to the highly individual and unpredictable reactions of the subjects.

This book presents an alternative framework for lexico-typological analysis that is applicable to a wide range of semantic domains, including those with a strong subjective component, e.g., pain predicates. We demonstrate the potential of this method by studying several domains of physical qualities: SHARP/BLUNT, FULL/EMPTY, WET, OLD, as well as dimensions, temperature, and surface texture.

---

1. This relativity of physical qualities is acknowledged from a slightly different angle in formal semantics. See the discussion by B. Partee of the striking example *a **tall** 14-year-old* vs. *a **tall** basketball player* (Partee 1995).

Most of the domains in our book (for example, WET, SHARP etc.) thus far have not been examined in a full-scale typological study. Meanwhile, the others, particularly the lexicon of dimensions – LONG, TALL, DEEP etc., – may appear to have been thoroughly studied (Bierwisch 1967; Bierwisch & Lang 1989; Lang 2002); nevertheless, the change in methodology we propose leads to important new observations about the structure of each domain. For instance, colexification (i.e. expression by the same word, cf. François 2008) of 'high' (as in 'high fence') and 'deep' (as in 'deep river') attested by us in some languages such as Komi, had not been encountered in the data of previous research, but can be easily accounted for within our framework (see Chapter 5).

2.   Methodology

The method adopted in this study primarily relies on lexical distribution. In doing this, we follow the rich lexicographic tradition of the Moscow Semantic School (cf. in particular Apresjan 1974, 2000), which has perfected the art of describing synonyms. Partial translational equivalents that form the material of lexical typology on a certain level are cross-linguistic near-synonyms which can be compared and contrasted in the same ways as a group of synonyms within one language. Subtle differences and nuances of meaning are revealed by finding contexts in which one term cannot be replaced by another. A well-known example of near-synonyms having different collocations within a single language is *wide* vs. *broad*, described by A. Wierzbicka (2006): *wide/*broad board*, but *broad/*wide back*. A similar effect can be observed across languages, cf. the co-occurrences of the English term *thick* and its equivalents *tolstyj* and *gros* in Russian and French respectively:

Table 1. *Thick* and its equivalents in Russian and French

	'finger'	'book'	'fog'
thick	thick finger	thick book	thick fog
tolstyj	tolstyj palec	tolstaja kniga	*tolstyj tuman
gros	doigt gros	*livre gros	*brume grosse

We assume that the cognitive reality behind these diagnostic contexts correlates with the prototypical situations that constitute a semantic domain (or "frames", in our terminology). The necessary prerequisite for considering a situation to be prototypical is that it must correspond with a lexicalized distinction in at least one language (cf. French *doigt **gros*** 'thick finger' vs. *livre **épais*** 'thick book' or Russian ***tolstaja** kniga* 'thick book' vs. ***gustoj** tuman* 'thick fog').

Regarding predicates, we follow Fillmore's view of a frame as a propositional form with slots for the situation's participants (cf. Fillmore 1976). Yet, an essential distinction made by lexical typology is that each slot in a frame is defined with respect to the semantic type of its filler. Qualities are particular in that the frames as basic situations are to a large extent determined by the kind of object this quality is ascribed to. Thus, the frames in the domain of thickness are represented, respectively, by 'finger', 'book', and 'fog' in the examples above.

A frame in the above sense is an extremely useful concept for lexical typology, since word meanings from multiple languages can be compared through a common inventory of frames in the domain in order to maintain compatibility of definitions and to highlight the various perspectives in which individual languages represent the reality. This assumption forms the groundwork for the theoretical approach advocated by the members of the Moscow Lexical Typological Group (MLexT).[2]

Minimal prototypical situations are used in MLexT to compile questionnaires; this is a practical implementation of our approach which offers an effective tool for facilitating routine fieldwork. These questionnaires are analogous to those used to identify grammatical categories, meanings, and clusters in a cross-linguistic perspective (see Comrie 1976, 1985; Dahl 1985 for TAM-categories, Corbett 1991 for gender and Corbett 2000 for number, etc.)

In a sense, our approach picks up the tools of grammatical typology and applies them to the lexicon. Frame-based questionnaires allow us to locate the distinctive features and oppositions which define the structure of a domain.

Another aspect in which our approach draws from the techniques of grammatical typology is the method of data visualization. We have adapted the traditional manually constructed semantic maps (see van der Auwera & Plungian 1998; Haspelmath 1997, 2003, cf. also Georgakopoulos & Polis 2018) to suit our needs. The maps allow us to illustrate the scope of lexical meanings in a domain, and to compare lexical systems graphically. Nodes on the map represent frames, and when two frames are connected, this means that they are colexified in some language. As soon as lexemes of individual languages have been plotted over this structure,[3]

---

2.   MLexT was started more than 10 years ago; it brought together young linguists from Moscow's state universities and academic institutes, as well as professional researchers in various languages and typology. All the authors of this volume are members of MLexT. For more information about the Group, see lextyp.org, for the detailed analysis of the theoretical background, see Rakhilina & Reznikova 2016.

3.   Throughout the volume, the term *semantic map* is used in two senses. First, it refers to a certain geometric configuration of frames that is hypothesized to hold cross-linguistically (the basic semantic map of a domain). Secondly, it may imply a result of plotting lexemes from a particular language on the basic map (the semantic map of a domain in language X).

one can instantly determine where colexification strategies coincide or differ; e.g., whether a given language colexifies blades and spikes in the domain of sharpness (i.e. 'a sharp knife' and 'a sharp spear'), and whether hollow and empty objects are colexified in the domain of emptiness ('a hollow gourd' and 'an empty cup') (for more details, see Rakhilina & Reznikova 2016: 107–113).[4]

Besides manual collection and processing of data, we make use of the latest advances in computational linguistics. Indeed, one of the challenges currently faced by lexical typology – as computational linguistics and corpora continue to make headway – is to suggest methods for automatic collection and preliminary analysis of data (for discussion of parallel corpora in cross-linguistic investigations, see Cysouw & Wälchli (eds) 2007). In particular, it is important to increasingly introduce statistical methods, since they impart greater weight to typological generalizations. In a separate section of the present volume we discuss distributional semantic models (cf. Baroni et al. 2014) and their applications to lexical combinations for the purpose of verifying and corroborating the choice of diagnostic contexts and main frames.

In its spirit, the frame approach we use is similar to that of the study on TEMPERATURE (Koptjevskaja-Tamm (ed) 2015) which was partly inspired by the earlier joint research (Koptjevskaja-Tamm & Rakhilina 1999, 2006). Our method easily uncovers the structure of this lexical domain. The primary opposition around which the domain is built involves the juxtaposition of whether the temperature of objects or their surroundings is experienced directly, by touch, or through the air. This is an instance of the well-known factor of anthropocentricity in language (cf., e.g., Wierzbicka 1988): the human body becomes a sort of a fixed universal reference point; this principle is manifested in the distribution of terms in the TEMPERATURE field where a distinct group for tactile experiences above and below the human body temperature, and a separate zone for atmospheric heat and cold are observed. For example, Russian has two terms in the subdomain of high temperature which are used in different contexts (meaning that they correspond to different frames) – *gorjačij* is used for the situations and objects that allow of tactile contact, while *žarkij* describes indirect experience, e.g., *gorjačij/*žarkij utjug* 'a hot iron', but *žarkoe/*gorjačee solnce* 'hot sun' (for a detailed account of the lexical oppositions in the domain of temperature, see Chapter 8 in this volume).

By taking collocation as a basis for our research, we step away from the mainstream methods in current lexical typology – namely, from the already mentioned

---

4.   It should be noted that lexico-typolgical semantic maps, unlike grammatical semantic maps, do not attempt to depict the stages of the diachronic development of meaning, and the connected nodes do not correspond to the historical senses of a term. Our maps represent the structure of a given fragment of the semantic space only in the synchronic perspective.

psycholinguistic experiments which are widely practiced by members of the Language and Cognition Group at the Max Planck Institute for Psycholinguistics (see, for instance, Ameka & Levinson (eds) 2007; Majid & Levinson (eds) 2011; Kopecka & Narasimhan (eds) 2012), as well as from introspection which is most prominent in the works of A. Wierzbicka and C. Goddard (Wierzbicka 1999, 2007; Goddard & Wierzbicka 2007; Goddard 2012) and in the early works of the Moscow Semantic School such as (Apresjan 1974, 2000).

The experimental method is most effective when describing objects or events that can be immediately perceived by an external observer, e.g., colors or postures. Situations with a strong subjective component, like pain or temperature, in contrast, are difficult to capture experimentally. The frame approach adopted by the members of the MLexT Group appears to be more powerful in this respect; it allows us to expand the scope of our research and include the lexical items that refer to highly subjective events or qualities, as well as words that carry a considerable evaluative component (cf. *soft* vs. *mushy* 'undesirably soft').

Conceptually, our approach is closer to Wierzbicka's school of thought, according to which the ideas of qualities are inherently anthropocentric, as they emerge from the ways in which people interact with objects (Goddard & Wierzbicka 2007). Therefore, a physically similar quality in a different object, or in a different context, is essentially a different quality, justifying the use of a separate term to describe it, as in the *soft*/*mushy* example. The data from all of the domains studied by the MLexT Group lends support to this assumption. However, methodologically, our approach differs from that of Wierzbicka and her colleagues. Their typology relies on a small vocabulary of semantic primes, hypothesized to be universal, which is reported to be sufficient for the purpose of expressing any meaning in any language; nonetheless, this technique appears to be less effective for larger groups of near-synonyms – and a comprehensive typological study unavoidably has to deal with a large number of terms. Moreover, the universal vocabulary of semantic primes is not designed to capture every opposition relevant to the structure of a semantic field.

We have tested the frame approach on several cognitive domains; some of them have been investigated with evidence from dozens of languages, namely AQUAMOTION, i.e., motion in a liquid medium (Maisak & Rakhilina (eds) 2007; Koptjevskaja-Tamm et al. 2010; Lander et al. 2012), PAIN (Britsyn et al. (eds) 2009; Reznikova et al. 2012), ROTATION (Kruglyakova 2010; Rakhilina 2010), FALLING (Rakhilina et al. (eds) 2020). The present book offers a collection of studies in which we apply the frame approach to the field of physical qualities.

3.   Language samples

Our language samples vary from domain to domain; taken together, they include Germanic (English, German, Norwegian, Swedish), Romance (French, Spanish, Italian), Slavic (Russian, Ukrainian, Byelorussian, Czech, Polish, Lower Sorbian, Upper Sorbian, Serbian, Croatian, Bulgarian, Macedonian), Baltic (Lithuanian, Latvian), Celtic (Welsh, Irish), Uralic (Finnish, Estonian, Hungarian, Khanty, Mari, Moksha, Komi, Udmurt, Nenets), Caucasian (Georgian, Besleney Kabardian, Aghul, Avar, Itsari Dargwa, Ingush, Lezgian, Tabasaran, Udi), Altaic (Azeri, Buryat, Kazakh, Kirghiz, Tatar, Turkish), Chukotko-Kamchatkan (Chukchi), Mayan (Ch'orti'), Kwa (Akebu) and Mande (Kla-Dan) languages, as well as Mandarin Chinese, Japanese, Korean, Mongolian, Malay, Hebrew and Basque. The data from some minor languages, like Uralic or Northwest Caucasian, is valuable per se, as it was collected by the authors from first-hand sources during field expeditions. Thus, in addition to our main task, this volume presents exclusive new material on lexicalization in endangered languages. No less important is the presence of a sign language in our sample along with oral languages – see Chapter 10 on the vocabulary of qualities in the Russian sign language.

Our sample may seem unbalanced and not diverse enough. By including so many closely related languages, do we not risk obscuring the overall typological picture? The answer is no, we do not; in the course of studying our sample we have arrived at an important theoretical observation: in our experience, an intragenetic sample is no less valid for lexical studies than a sample of unrelated languages.

As a matter of fact, relevance of intragenetic typology has been noted even by some grammarians (compare Kibrik 1998, 2009 on intragenetic typology in grammar). While this idea may be disputed in relation to grammatical typology, we consider that for lexical typology it has proved relevant almost beyond doubt. Our previous studies of a variety of domains, from swimming to swaying and rotation, had repeatedly testified that the lexical systems of genetically related languages are no more likely to be similar to each other than to a completely unrelated language (cf. also Majid et al. 2007). The same is true when we examine the domains of qualities discussed in this volume. For example, in the domain of surface texture, the data from the 10 Uralic languages was found to demonstrate the same semantic oppositions and metaphoric patterns as the five languages from three other families (see Chapter 6).

This is not a coincidence, but a consequence of the fact that vocabulary is subject to quicker change than grammar. As noted in our earlier work (Rakhilina & Reznikova 2016), while grammatical constructions take centuries to evolve, vocabulary is much more fluid. A single generation of speakers may witness words falling in and out of use and word meanings changing dramatically. Consequently,

even a relatively short – from a historical point of view – interval can be enough for related languages to form very dissimilar lexical structures and to develop some typologically relevant distinctions.

4. Data selection

The chapters of this volume present seven domains, examined either individually or paired with their respective antonyms. These are: the property words from the fields of dimensions, temperature, and surface texture, as well as the qualities SHARP/BLUNT, FULL/EMPTY, OLD, and WET. Besides them, our research spans other physical qualities such as DENSE/SPARSE, TIGHT/LOOSE, HARD/SOFT, HEAVY/LIGHT, VISCOUS and CLEAN/DIRTY. Due to the limitations of this introductory chapter, we will not focus on each of these domains in detail here; however, we will be touching on them in the subsequent sections of the Introduction when discussing the general characteristics of the terms of qualities.

Typically, a domain in a given language is represented by several lexical items distributed over the frames which constitute the domain. For example, the domain of wetness in English is covered by the adjectives *wet, damp, humid, moist* and *dank*, in Khanty – by *jinki, jinkišək, ńar* and *nivəŋ*, and by *šlapias* and *drėgnas* in Lithuanian. However, it is often the case that a term is not limited to its own domain – it also embraces some frames which clearly belong to a different semantic field. This can be observed, for example, in one of the dimensional subdomains which we presented in Section 2 in order to illustrate the notion of frames: in some languages, including English, the terms for measuring layers (*thick* or *thin* books, blankets, ice, etc.) can also be used with substances (porridge, fog etc.), cf. also Georgian *skeli çigni* 'thick book' and *skeli supi* 'thick soup'. What *thick* and *thin* denote in such collocations is not measurements – because substances are amorphous – but consistency, a completely different physical characteristic. Other languages use a dedicated term for thickness of a substance, cf. Mandarin **nóng** *tāng* 'thick soup' vs. **hòu** *shū* 'thick book'. Still others describe thick substances with the same terms as tightly-packed sets, i.e. 'dense', such as Russian **gustoj** *sup* 'thick soup', **gustoj** *les* 'dense forest' vs. **tolstaja** *kniga* 'thick book'.

A similar phenomenon occurs when a word denoting 'heavy' is used for stiff pedals or buttons. At first glance, the situations behind 'heavy' and 'stiff' seem unrelated – heaviness implies an experience of lifting the object, while a stiff button is pushed, not lifted. Yet in Japanese, for example, the same term can be used in both situations: **omoi** *baggu* 'a heavy bag', **omoi** *botan / pedaru* 'a stiff button / pedal'. This can be explained either as a metonymy, since a stiff button or pedal requires the operator to put their weight into pushing, thus acting similarly to a heavy object, – or

as a metaphor, since the mechanism resists being pushed and requires effort to operate it, similarly to how a heavy object requires effort and resists being lifted.

It must be noted that pedals or buttons are not the prototypical objects of the English adjective *stiff* either. This term primarily describes objects that are difficult to bend: *stiff leather, a stiff back* etc.; however, a pushing action can be used to test the stiffness of some objects from outside this group, typically those made of hairs – *a stiff brush, stiff bristles*. We could therefore speculate that this action creates a bridge between the two situations, allowing the adjective *stiff* to extend over to unyielding mechanisms.

As we see, it takes certain inventiveness and a cognitive effort to come up with the idea to apply the terms for 'heavy' or 'stiff' to buttons and pedals. An easier, more natural transition might be expected towards the words that already characterize their objects as resisting pressure – that is, the lexemes with the meaning of 'hard'. Indeed, it is the terms of hardness that the majority of languages in our sample use for stiff mechanisms, e.g., Italian *duro* (cf. *un pedale duro* 'a stiff pedal') or Serbian *tvrd* (cf. *tvrdo dugme* 'stiff button'). Even in Japanese the adjective *katai* 'hard' is found in such contexts, as an alternative to *omoi* 'heavy': e.g., *katai botan* 'a stiff button'.

We quote these examples here to illustrate an important feature of how the whole realm of qualities is structured. The examples show that languages differ not only in the scope of terms within a particular domain and the oppositions they form; there is also significant variation as to where the boundaries of the domain are drawn. 'Thick substance' can be grouped with the terms of dimensions or denseness, while 'stiff mechanism' can be placed under the terms of hardness or heaviness; such "liminal" frames function as links between domains. Their existence suggests that qualities flow into one another, forming a single continuum rather than a set of neatly delineated semantic fields. The time-honored scientific metaphor of a domain as a separate entity, so familiar to lexicologists, turns out to be an illusion: lexical typology finds no natural boundaries between domains.

The view of a semantic field as an entity which is isolated, self-contained and reducible to a limited set of differential features is a legacy of structuralism. The question of establishing synchronic, not necessarily metaphoric, connections between domains has not been systematically raised before. Yet in reality some domains appear to be joined together via common frames.

Stepping outside of a wide-scale typological study, we may find further evidence in support of this interconnectedness of semantic fields; it is corroborated on the morphological level by the structure of disyllabic compounds in Mandarin Chinese. Most Mandarin compounds of qualities consist of two components, each representing a quality, so that the components are related but not redundant, such as *róuruǎn* 'soft', lit. 'flexible-soft', or *jiānyìng* 'hard' lit. 'strong-hard'. Such pairs can be said to embody the links between the adjacent domains. We have found that

the meanings involved in Mandarin disyllabic compounds tend to overlap from a typological perspective, i.e. they reveal common frames. For example, the connection between the domains of dimensions and denseness discussed above is echoed by the Mandarin compound *nónghòu* 'thick (about air)', lit. 'dense-thick (in size)'.

Consequently, it must be taken into consideration that the seven domains addressed by the respective chapters of this volume have no clear boundaries either, and their scope (as outlined in the chapters) is largely a matter of convenience. Let us now take a closer look at the inner structure of the domains.

5. Lexical oppositions in qualities

Basically, each domain is formed around its own set of lexical oppositions, but a challenging question arises: do these sets have any common features, or, in other words, is a general typology of qualities possible? If such a common basis for lexical variation exists, it would considerably facilitate the research of any given domain of a physical quality, since we would know beforehand what kinds of oppositions to look for. This section gives an overview of some of the oppositions that are relevant across the domains, which we have discovered so far.

We find that the recurring patterns which motivate the diversity of qualitative terms in a given domain tend to be determined by the following three factors: (a) properties of the objects which are being described; (b) the number of semantic arguments, because not all qualities are one place predicates; (c) characteristics of the quality itself.

The first factor concerns the semantic arguments of the qualitative predicate, generally, *taxonomy* (see 5.1) and *topology* (5.2). The second concerns the additional lexical variety which arises when the qualitative term can have *additional arguments* (5.3). The third factor reflects the anthropocentric context that is typical of the conceptualization of qualities – primarily the *evaluative component* (5.4) and the *channel of perception* (5.5).

5.1   Taxonomy and mereology

The basic taxonomic opposition observed in a wide variety of qualities is **animacy vs. inanimacy** of the head noun, i.e. of the main participant. The role animacy (or sometimes humanity) of the main participant plays in lexicalization is well known for predicates: the choice of a term is often determined by whether its subject is human or non-human/inanimate. This distinction can be observed, for example, in the domain of motion in liquids in such oppositions as with English *swim* for animate

agents vs. *drift* for inert inanimate objects. If the situation itself implies animacy, the distribution may be between human and animal agents, see the German verbs of eating and drinking: *essen / trinken* 'eat / drink (about humans)' vs. *fressen / saufen* 'eat / drink (about animals, when used in their direct meaning)', or the Russian verbs of dying: *umeret'* 'die (about humans)' vs. *izdoxnut'* 'die (about animals)'.

A similar phenomenon can be observed in property words. Certain qualities are described by different terms, one for people and another for inanimate objects.[5] Thus, both people and things have physical height, but German tends to use mostly the adjective *hoch* for tall objects and *groß* for tall people. Dedicated color terms for hair and eyes such as English *blond* or Russian *karij* 'brown (about eyes)' are another good example for anthropocentric qualities.

This effect is so widespread that for every property word that can be used in the same sense to describe both humans and non-animate objects, we may expect some other language to have a corresponding pair of terms. In the domain of oldness English has the general term *old*, which covers both people and artifacts (*old men, old coins, old shoes,* etc.), and looking at the German *alt*, French *vieux*, Russian *staryj,* with their similar scope of meaning seems to suggest that this type of colexification is universal, especially since people and artifacts are equally affected by the passage of time. Still, we do find languages like Japanese, which makes the distinction between animate and inanimate objects (*furui uwagi* 'an old coat' vs. *oita inu* 'an old dog'), or Georgian with a specific term for humans, *moxuci*, vs. *dzveli* for everything else. Georgian also has a special term for old plants (*beberi*), which can be applied to people in pejorative contexts.

The anthropocentric nature of language can be observed with equal clarity in the opposition of **natural vs. artificial objects**, as in the two Russian terms for multiple uniform elements located closely to one another: *gustoj* (*gustoj les* 'a dense forest', *gustye volosy* 'thick hair') vs. *častyj* (*častyj greben'* 'a fine-toothed comb', *častaja setka* 'fine mesh').

Beside such general oppositions, of course, there are more specific ones limited to a single domain. For example, transparency or opaqueness can be expressed differently depending on whether the object is a solid, a liquid or a gas – English has a specific term for solids vs. gases and liquids (**opaque** *glass* vs. **turbid** *water/air*), Mandarin – for liquids vs. gases and solids (**húnzhuó** *de chítáng* 'a muddy pond' vs. **bùtòumíng** *de kōngzhōng* 'turbid air').

---

5.   While some qualities apply to both persons and objects, it must also be noted that there exists a large class of qualities specific to people (e.g., 'clever' or 'proud') that can only be applied to objects or actions metonymically, as well as qualities specific to objects (e.g., 'bitter' or 'tough') that apply to people metaphorically.

An even more specific opposition can be seen in the domain of hardness. This domain often contains terms for particular kinds of food, especially meat and bread, like the Russian *čerstvyj* 'stale (about bread)' vs. the general adjective *žestkij*, which covers a wide variety of other objects (*žestkaja poduška* 'a hard pillow'*, žestkoe mjaso* 'tough meat'), or the Finnish *sitkeä* 'tough (about meat)', also opposed to the broader adjective *kova* 'hard, tough': *kova sänky* 'a hard bed', *kova pinta* 'a hard surface', *sitkeä liha* /*[??] kova liha* 'tough meat'.

At present it is difficult to judge the value of such domain-specific distinctions. More domains, and not only of physical qualities, need to be studied before any reliable conclusions can be made.

On the whole, the taxonomic features, as they appear in the lexicon, run through the whole lexical system of a language, irrespective of parts of speech, as with animacy which, as we have seen, is relevant for verbs as well as adjectives. The same is true for mereological features. We have already seen above that in the domain of thickness separate terms may exist for objects ('a thick book') vs. masses ('thick fog'); this is a taxonomic distinction. Similarly, a term may be restricted to sets of objects – this is when the distinction from an individual object is mereological. For verbs it can be observed in the domain of falling, where all three are often lexicalized, as in Besermyan Udmurt: *kiškanə* 'fall/pour' (for masses such as sand), *uš'ənə* 'fall' (for individual objects such as an apple), and *kərdəne* 'fall' (for sets, such as a tray of glasses).

For quality words, too, an object's mereological properties may be relevant when different terms exist for a single item vs. a collection, or a member of a set. Russian uses the adjective *bol'šoj* 'big' for individual objects and its near-synonym *krupnyj* 'big' for objects viewed as representatives of their kind: **bol'šoj** *dom* 'a big house' vs. **krupnye** *kapli* 'big raindrops', **krupnyj** *zajac* 'a large hare'. There is also a typological tendency for sharpness to be expressed by different terms when talking about a single spike vs. a multitude of them or a spiky surface, as in German: *spitzer (*stacheliger) Dorn* 'a sharp thorn' vs. *stacheliger (*spitzer) Busch* 'a thorny bush'.

There is a traditional view of mereological oppositions in the qualitative vocabulary as a basis for metonymy (see, e.g., Waag 1901; Apresjan 1974, 2000; Radden & Kövecses 1999; Peirsman & Geeraerts 2006) – see such classic examples as *a **merry** youth / company, a **rich** farmer / village*. However, we see that they can influence lexical choice directly. This rises an interesting question: what kinds of collective objects will typologically tend to be described by the same quality words as their elements, and what kinds will have their own qualifiers. Preliminary data suggest that groups of people, compared to collections of objects, are more often conceptualized as inheriting the properties of their members through metonymy. In other words, "animate" collectives tend to share vocabulary with their individual

members, and in this sense resemble their members, while the static "inanimate" collections are ascribed completely different properties. On the whole, though, this question requires separate typological investigation.

5.2   Topology

An object's topology, i.e. its characteristic shape, size and position, is another factor that can influence the way this object is described (cf. Talmy 1983; Rakhilina 2000). It is particularly clear in the domain of dimensions, and Chapter 5 provides numerous examples of lexical choice relying on shape and orientation. Let us give just one of them here. Terms for height tend to divide objects into classes based on shape: barriers such as walls or fences, which are both tall and long, and poles such as trees or towers, for which height exceeds all other dimensions; such distribution can be seen in English, cf. *a **high** fence/wall* vs. *a **tall** tree/candle/skyscraper/mast.* Kazakh is somewhat similar: it uses the adjective *biìk* 'high' for both poles and barriers, and has *uzyn* specifically for tall poles. In the subdomain of low height the opposition of poles and barriers can be observed, for example, in some Nakh-Daghestanian languages such as Aghul, which describes low barriers as *χeba* and low poles as *džiqqe* (see Chapter 5 for details).

Lexical oppositions based on shape are not limited to the dimensional domain. The classes of blades vs. spikes form a major opposition in the domain of sharpness, that can be found in a significant portion of our sample, e.g., Hungarian *eles* vs. *hegyes*, and similar pairs, such as Komi *lečyd* vs. *jues*, Mandarin *fengli* vs. *jiān*, Besleney Kabardian *ž'an* vs. *pamçe* (see Chapter 2).

A topological interpretation is also possible for the opposition of containers such as cups, boxes and cupboards, and other objects that may possess a cavity, such as pipes, gourds or shells. This distinction tends to be lexicalized in the domain of emptiness (see Chapter 4). However, function plays at least as large a part in it as shape. It has been repeatedly pointed out for various domains (see Wierzbicka 1985; Rakhilina 2000) that the way objects are described is usually determined by how people use them. Wierzbicka observes, for example, that the properties of being 'made to drink hot liquids from' and 'small enough for people to be able to raise them easily to the mouth with one hand' constitute an essential part of the concept of a cup (Wierzbicka 1985: 59). Similarly, function is central to the opposition of empty and hollow (or full and solid) objects: only containers proper, intended to hold something, can be described as full or empty, while non-functional objects are characterized as hollow or solid based on their shape. Function in this case takes precedence over the topological opposition.

Function can also completely override an object's topological properties. It is remarkable that an example of this can be found in the seemingly impersonal domain of dimensions. Komi has several terms for height of vertical objects, in particular, 'tall' trees and 'high' fences are qualified with different adjectives. The adjective for fence-height is also used when speaking about 'deep' rivers. This effect can be accounted for if we consider these objects from the point of view of someone who finds them in their path: both high fences and deep rivers are obstacles that are difficult to cross, whereas standing trees are not usually regarded as an obstacle. This pair of terms illustrates how lexical distribution may ignore topological similarity (the vertical orientation, common to fences and trees) and give priority to the way humans interact with their surroundings.

5.3   Number of arguments

Studies of verbs show that all semantic arguments, and not just the subject, may influence the lexical diversity in a given domain. This means that the more participants a situation implies, the more typologically varied its lexicalization will be. Indeed, if the lexical choice for a one-place predicate necessarily relies on the properties of its only argument (e.g., as we have discussed, separate terms often exist for people, animals and inanimate objects), every additional participant contributes to the range of lexical variation.

The verbs of eating and drinking tend to differentiate between human vs. animal agents and different kinds of object; one common opposition is that of solid and liquid food, but sometimes a verb may imply even more specific objects, cf. a special word in Russian for eating something that tastes delicious *lakomit'sja*. The lexical choice may also be influenced by whether the subject uses cutlery or eats with his/her hands.

A still richer domain is that of cutting and breaking (cf. Majid & Bowerman (eds) 2007). There are not less than four participants which can influence lexical choice (see Kashkin 2010 for details): the agent, the patient, the instrument and the result. The opposition relevant to the first participant is whether there must be an agent in control of the situation, or the subject can be a natural force such as wind, hale or lightning. In Erzya *pol'enč'ams* 'to chop (wood)' is always a purposeful action, while the verb *tapams* 'to break' describes both actions and uncontrolled events – e.g., both the breaking of a glass hothouse caused by hail, and by someone hitting it with a hammer. For the patient texture is often an important factor. Erzya again has two verbs of cutting with a sharp instrument - *pečkems*, reserved for relatively soft objects and substances like meat or paper, and *ker'ams*, which can also apply to hard materials like glass or bone. Another relevant property is the object's topology. Tearing by hand in Komi is described by *orədny* for long, thin

rope-like objects and by *kos'avny* for flat objects such as a sheet of paper or cloth. Many languages lexicalize a particular kind of instrument. For example, in English *hewing* usually implies an axe and *cutting* – a blade or a saw, while Russian has the predicate *pilit'* specifically for sawing. Result, namely the size and number of the resulting fragments, may also form the basis of a lexical opposition (cf., e.g., the German verb *zertrümmern* which implies multiple pieces as result of the action).

It must be made clear that we list the arguments and their types like this only to classify and organize the vast variety of contexts in which, as the corpora show, the predicates of cutting and breaking can be found. In other words, we do not take the structuralistic view of properties as independent semantic components, and do not assume that enumerating their combinatory possibilities would give us the universal picture of the domain (compare Lehrer 1974, a study of cooking verbs that does rely on this paradigm).

Lexicalization in a wide variety of domains demonstrates that properties are largely interconnected, and often selecting the value of one property either determines the value of another or makes it irrelevant to the situation. Thus, for the verbs of eating, if food receives a positive evaluation, the prototypical agent becomes limited to people, and the presence or absence of instruments irrelevant. We encountered no language with two separate terms for enjoying food, one with cutlery and another with one's hands. Values do not co-occur arbitrarily; rather, they form holistic gestalts. In our belief, lexical typology must be able to provide an answer to which gestalts can be lexicalized by dedicated terms (what we call frames), and which tend to be combined with others.

Going back to the qualitative predicates, most of them have only one argument, but for the rare two-place predicates the second argument increases the possibilities for lexical variation by introducing additional taxonomic and topological oppositions. For example, the situation of fullness/emptiness implies two participants, namely the container and contents, and either of these may influence lexical choice. Chapter 4 gives examples of adjectives reflecting the topology of the container (differentiating between containers proper, hooks, surfaces, locations, etc.), as well as the taxonomy of the contents. The latter case is found in Serbian: places are described by different terms depending on whether they are empty of people, when the adjective *pust* is used, or things, when the preferred adjective is *prazan*. This opposition is beginning to fall out of use in modern Serbian, but what is crucial in this example is the very fact of its existence in a natural language.

Wetness is another domain where the second semantic argument affects the choice of the qualitative predicate. While the first argument is the wet object itself, the second is the source of moisture. There, dampness caused by humid air may be contrasted with direct contact with a liquid, as in Georgian: *nest'iani* 'damp from atmospheric moisture' vs. *namiani* 'damp, e.g., after washing'. Alternatively, specific

kinds of liquid may be relevant, as in Hungarian which has a dedicated adjective *vizes* 'wet with water' in addition to the more general *nedves* 'wet'.

Above we have concentrated on the arguments, namely on their number, taxonomic and topological properties. Let us now see how the interpretation of the quality itself can be coded by a specific term.

5.4   Evaluation

When using an object for some purpose, a person perceives its properties and evaluates them as either positive (i.e. suiting the process) or negative (hampering it). This evaluation can become lexicalized. In this case, typically, there is one term for a neutral description of the quality in question, and another with added positive or negative connotation. Thus, in the domain of softness English distinguishes the neutral *soft* from the negative *mushy* that covers excessive, undesirable softness, e.g., of overripe fruit. One of the Dogon languages, Tomo kã, has the neutral term *hwèjí-hwèjí* and the positive one *bùrè-bùrè,* which is limited to pleasant softness, like a soft cushion or grass. A similar situation can be observed in French with its adjective for 'pleasant, cozy, soft touch', *moelleux*, as in *fauteuil / coussin / peignoir moelleux* 'a soft armchair / cushion / robe'. The specific of the French system is that the other term in the opposition, *mou* 'soft', is not always neutral (though it does allow such usage, as in *sable mou* 'soft sand' or *cire molle* 'soft wax') and can carry negative connotation, as in *pain mou* 'soft/soggy bread', *lit mou* '(too) soft bed'.

There are other domains where lexicalization is influenced by evaluation. The domain of wetness tends to have a dedicated negative term; indeed, moisture hugely impairs the function of some objects – a damp cellar is bad for storage, damp matches won't light, etc. (see Chapter 3 for details).

Neutral and negative terms also coexist in some languages for 'viscous/sticky' objects and substances. In Russian a surface can be described as *klejkij* when that is supposed to be its normal state, as in *klejkaja lenta* 'adhesive tape', or as *lipkij* when the state is abnormal and undesirable, as in *lipkie ruki* 'sticky hands' (from handling something).

5.5   Type of perception

Lexical diversity within a domain may be motivated by the fact that qualities can be perceived through different channels – thus, we can touch an object to find whether it is wet, but sometimes a look is sufficient.

A related phenomenon, commonly known as cross-modal metaphor, is observed in terms that have several senses which correspond to different sensory

channels, such as *loud* in *loud sounds* (auditory) and *loud colors* (visual). This phenomenon may receive several theoretical interpretations, one of which is treating it as a metaphor (cf. Shayan et al. 2011; Levinson & Majid 2014; Speed et al. (eds) 2019). This approach, however, has a weak point: the source and target of the shift between the two senses of the word would then lie on the same level of abstraction, since both describe perception, while classic metaphoric shifts are supposed to run from concrete to abstract. Our approach differs from some of the other lexico-typological studies in that we do look into metaphoric shifts. However, we consider them separate from cross-modal perception – we would say that *hot* in *hot asphalt* (tactile) is connected to *hot argument* or *hot news* by metaphoric shifts, with a step up in abstraction, and to *hot curry* (gustatory) or *hot pink* (visual), which describe similar sensory experiences, by cross-modal shifts (cf. the notion of intrafield extensions in Viberg 1984; Evans & Wilkins 2000). We treat the latter as an instance of metonymy, where the focus of attention shifts between channels of perception, see Ryzhova et al. 2019.

Lexical typology brings into focus the inverse situation, when a similar experience received through different sensory channels is expressed by separate terms. It is often encountered when comparing languages; so the mode of perception is not only the basis of polysemy for some qualitative words, but also one of the sources of lexical oppositions between them. One such case is surface texture. A surface with no irregularities can either be experienced by touch, like skin or a tabletop, or by sight – land or walls do not require tactile contact to determine that their surface is even. This opposition is lexicalized in Estonian which differentiates between *sile* 'smooth to the touch', as in *sile nahk* 'smooth skin', *siledad juuksed* 'sleek hair', and *tasane* 'visually flat', as in *tasane põrand* 'a smooth floor', *tasane karjamaa* 'flat pasture' (See Chapter 6 for detail).

Another example of lexical differentiation between tactile and visual experience can be seen in the domain of softness. Korean has the term *pwutulep-ta* for cloth, paper, toothbrush, dough etc. – the functioning of all of them involves touch. People wear cloth on their body, hold and crumble paper, brush their teeth with toothbrushes and knead dough, therefore directly experiencing the softness of these objects and substances. The opposing term, *mongsilmongsil-ha-ta*, describes hypothetical softness that one assumes from how an object looks. We found examples where this term was used about sheep's wool, cotton candy and body parts (cheeks and hands), and all of them involved seeing, but not touching.

The domain of sharpness gives a third example of sensory channel playing a part in lexicalization, again distinguishing touch and vision. Functional sharpness, characteristic of cutting and piercing tools, is associated with a tactile experience. On the other hand, for the objects which are shaped like spikes but not used for piercing, this property is experienced only visually. Chapter 2 shows that functional

and visual sharpness is often described by separate terms, e.g., English *sharp* (functional) vs. *pointed* (visual), Japanese *surudoi* 'sharp' vs. *togatta* 'pointed' (*surudoi kiba* 'sharp fangs' vs. *togatta enpitsu* 'a sharp pencil' – pencils are sharpened for use, but it does not involve piercing or stabbing).

One might claim that such cases speak against treating the cross-modal shifts like *sharp claw → sharp nose* as metonymy. The classic metonymy is expected to be so cognitively transparent and predictable as to repeat itself in almost any language. Indeed, the lists of metonymic shifts compiled for different languages support this belief by having a large overlap (see Apresjan 1974, 2000; Radden & Kövecses 1999; Peirsman & Geeraerts 2006). Conversely, if a pair of senses is expressed by different terms in a large number of languages, it may suggest that their colexification is not a real metonymy.

A closer cross-linguistic look, however, shows that the metonymic shifts are not as uniformly reproducible as they seem. Even for the most venerable of them, such as "author ↔ text", a usage that is possible in one language may be infelicitous in another – e.g., *The paper argues that…* is an acceptable English expression, while its literal translation into Russian ˀ*Statʹja utverždaet, čto…* is awkward at best. While a typological study of metonymy is a task for the future, the mere fact that the difference in sensory channel can motivate a lexical opposition should by no means preclude us from ranking cross-modal shifts as metonymy rather than metaphor.

6. Metaphoric shifts

The fact that we rely on lexical co-occurrence and diagnostic contexts also enables us to observe not only cross-modal metonymies, but also metaphors. This is an important contribution to lexical typology, because the metaphoric senses of quality words tend to be underrepresented in typological studies. The denotation-based empirical method is ill-suited to dealing with abstract entities which are the typical goals of metaphoric shifts, and cannot easily capture these kinds of meanings. Some studies of metaphors have been done for body parts (e.g., Wilkins 1996; Koch 2008) and predicates of cognition (Sweetser 1990; Evans & Wilkins 2000; Vanhove 2008), but not for property words. A few observations on the qualitative terms can be found in François 2008 (the typical metaphors for *straight* and its French equivalent *droit*) and in Goddard & Wierzbicka 2007 (metaphors of the English and Polish words for 'heavy' and 'sharp'), but they are made on a small number of languages and very limited data. The first study draws its data from dictionaries, the second is based on the authors' introspection, and neither can reflect the whole system of metaphoric shifts in the domains they touch upon. The typological study of metaphors led by Anna A. Zalizniak (Zalizniak 2008; Zalizniak et al. 2012; see also the

DatSemShifts database http://datsemshift.ru/) also takes its data from dictionaries and does not aim at an exhaustive description of individual domains.

The lack of research into qualitative words is also evident from Heine & Kuteva's (2002) dictionary; like Anna Zalizniak's database, it is a catalogue of semantic shifts, but of a particular kind – the shifts from lexical to grammatical meaning. Among the many sources of grammaticalization it only mentions three adjectives: 'bad', 'true' and 'alone'. Yet a closer look at the semantic evolution of physical qualities shows that grammaticalization of adjectives is much more widespread than that. Even our data offers a number of cases where content words move into the scope of grammar: there are adjectives that turn into intensifiers, like Mandarin *jiānruì de zhēn* 'a sharp needle' → *jiānruì de téngtòng* 'a sharp pain' or Russian *polnyj mešok* 'a full sack' → *polnyj proval* 'complete failure', and others that produce discourse markers, like English *purely, clearly, simply*, and others. The second scenario of grammaticalization is explored based on the adjectives of straight shape in Luchina et al. 2013; Luchina & Nanij 2016.

Semantic shifts need cataloguing and understanding, but their study has additional value. According to our theory and empirical observations (see Rakhilina & Reznikova 2016), metaphors can be used as a tool to distinguish frames. Different metaphorical meanings tend to go back to different physical frames. For example, only functional sharpness (as in *sharp knife/needle*), and not pointed shape (as in *sharp nose/roof*), gives rise to intensifiers, supporting the hypothesis that the two situations are cognitively distinct, cf. also the analysis of 'heavy' and 'weighty' in Ryzhova et al. 2019. This makes the examination of metaphors an important step, specific to our methodology, in verifying the results of a lexico-typological study.

We believe that a semantic shift starts with isolating and reinterpreting particular aspects of a given physical situation (frame). If we look at the domain of tightness, for example, we find the frame 'tight clothes' which describes insufficient space between clothing and the wearer's body. Such clothes are uncomfortable and constrict movement; this idea of discomfort motivates a typologically regular metaphor of shortage of resources such as time or money (Eng. *tight budget, tight schedule,* Fr. *budget serré* 'tight budget', *délais serrés* 'tight deadlines'). On the other hand, the frame 'tight rope' describes maximum tension; the stretched object risks breaking under the strain. This image becomes reinterpreted as a charged, unpredictable state of things that can go out of control at any moment, or a relationship on the point of breakup (Spanish *situación tensa* 'tense situation', German *gespannte Beziehungen* 'tense relationship'). The two frames represent different concepts which lend themselves to different shifts. We have observed this mechanism in a number of domains, qualitative (see, e.g., Chapter 6 of this volume) as well as predicative (see, e.g., Rakhilina 2007 on the metaphors in verbs of swimming and floating).

This does not mean that separate frames can never produce similar metaphors. In the domain of fullness 'full containers' and 'full sets' can be colexicalized, but some languages have different terms for them (English *full* vs. *complete*, Spanish *lleno* vs. *completo*), allowing us to view the two situations, fullness and completeness, as distinct. Nevertheless, both kinds of terms can evolve into intensifiers (Spanish *la banda sonora completa* 'the complete soundtrack' / *una victoria completa* 'a complete victory', Russian *polnyj stakan* 'a full glass' / *polnyj bred* 'utter nonsense'). In its counterpart domain of emptiness, there are examples of terms for 'hollow non-container objects' and 'empty containers' similarly describing speech or text as 'devoid of meaningful content' (see Chapter 4 for details).

We do not see these cases as contradicting the general tendency. Rather, the metaphors are similar because the underlying physical situations possess substantial similarity which the metaphors happen to highlight. Complete sets and full containers have a common feature: nothing can be added to them, they have reached the limit of their capacity. This upper limit translates naturally into the abstract idea of maximum. Similarly, emptiness and hollowness both imply the absence of an expected inner part or content, and the image can be applied to information as an abstract constituent or content.

Consequently, we hold to our claim that semantic shifts reflect the physical senses. Diverging metaphors remain a valid additional indication of the frame structure of a domain, while identical metaphors are motivated by the similarities in the original frames.

7.   Conclusion

This chapter gave an overview of the approach the Moscow Lexical Typological group takes when working with linguistic data, the key instruments we use for cross-linguistic analysis, and the kind of results – namely, the typologically relevant lexical oppositions – that our method produces for qualitative domains. The following chapters will show our methods in action. Each chapter showcases an aspect of our approach while describing the domain that is most suited to illustrate it.

We begin by defining frames and the procedure for their identification using the simplest data – the domains of sharpness and bluntness (Chapter 2). The authors analyze the terms for sharpness and bluntness in attributive contexts and give a step by step account of their method in minute detail. For both qualities they build a set of frames, organize them into a lexico-semantic map and demonstrate how the colexification strategies adopted by various languages are motivated and allow for a natural interpretation.

Sometimes the data is less transparent. In the domain of wetness (Chapter 3), the mere combination of the adjectival attribute plus its head noun is not enough to gain insight into this quality's semantics. The chapter demonstrates that the broader situation and, consequently, broader phrasal context needs to be studied if we want to clarify the system of oppositions behind the vocabulary of wetness. Chapter 3 also shows that the speaker's evaluation of the situation may affect the lexicalization of a quality.

The domain of fullness and emptiness, whose typological analysis is presented in Chapter 4, serves as an example of the fact that a quality can have multiple semantic arguments, in which case each of these, or their combined influence, can contribute to the lexical variety in the domain.

While additional arguments are a less common motivation for lexical choice, the next two chapters bring us back to factors more typical for qualitative predicates. The domain of dimensions, discussed in Chapter 5, is structured around shape and orientation. These two characteristics give rise to a whole system of topological oppositions that we use to describe the ways in which languages measure containers, rods, poles, belts and other topological classes of objects. Chapter 6 is devoted to words of surface texture (roughness, flatness and smoothness). One of the main sources of lexical oppositions here is the channel of perception – this domain tends to distinguish tactile and visual experiences.

A separate and equally important topic is the MLexT approach to metaphor. Individual metaphors are analyzed in detail in Chapters 2, 4 and 6. It can be seen that while there is no place for metaphors on the semantic map itself, derivative meanings still maintain a connection with their motivating physical senses and fall into the same structure of frames which constitutes the domain, i.e. metaphoric shifts take particular frames as their starting point.

In this way, Chapters 2 to 6 explain and illustrate our methodology: the notion of frames, systems of frames, and the principles for building semantic maps. We also discuss the basic factors that govern the oppositions within qualitative vocabulary: these factors define the frame structure for the terms of physical qualities and determine their potential for metaphoric derivation.

The remaining chapters deal in various ways with our language sample. The problem which is most often pointed out by our colleagues, and of which we are acutely aware, is that every time the sample appears to be too small and far from balanced. The reason is that, for every domain in Chapters 2 to 6, the data was collected and examined manually. This means a lot of painstaking, time-consuming work for the researchers, experts and native speakers involved, which is why we were only able to cover relatively few languages for each of the domains. Chapter 7 presents an experimental study of the domain of oldness, where dictionaries are

used as the source of data (whereas the detailed frame structure of the domain was initially constructed with our routine methodology). The vast lexicographic data was used to build semantic maps, and, on the whole, turned out to support our initial hypothesis as to the oppositions relevant to this domain.

Chapter 8 discusses the problem of using data from related languages. This is a major issue for us, as has been stated above: we have repeatedly claimed that related languages provide adequate material for lexical typology. Now, terms referring to temperature are used to prove that our claim also holds for qualitative words, since related languages demonstrate independent semantic developments, and the data from them perfectly fits into the general typological picture (in this respect, see also Chapter 6).

What is more, this picture continues to persist when we include a genetically and geographically distant language such as Mandarin Chinese. Chapter 9 shows that Mandarin, despite the non-trivial nature of its vocabulary where disyllabic and monosyllabic words exist in complicated relationships of near-synonymy, still adheres to the same basic oppositions that were previously found in the domains of surface texture, hardness and others.

Chapter 10 presents even more "exotic" data – it is a study of qualitative predicates in Russian Sign Language. The fact that this language is non-auditory makes surprisingly little difference for the structure of qualitative domains, which allows for two possible explanations. One is that the frames and oppositions which we have uncovered may have some kind of deep-rooted cognitive motivation, and then we may expect to find them in a wide array of world languages. Alternatively, sign languages may not always be as "exotic" as the visual-manual modality they use seems to suggest, and their semantics may be primarily determined by local culture. The latter explanation, however, does not hold in the case of Russian, as the structure of qualitative domains in Russian Sign Language drastically differs from that of spoken Russian. We may hope therefore that the first hypothesis is closer to the truth, and our results would still be valid for a larger sample.

Although even a small sample already allows us to predict the general structure of a semantic domain, in the future it still seems desirable to automate as much of raw data collection and preprocessing as possible; this would help us cover more languages and make our approach more universally applicable. The concluding chapter, Chapter 11, discusses our steps in that direction and the long-term goals and perspectives of the MLexT group.

It turned out that this book took quite a while to be written and published. In the meantime, its authors had 6 babies and defended 5 dissertations. At different stages the project was kindly supported, contributed and commented by our friends and colleagues: Valentina Apresjan, Yulia Badryzlova, Östen Dahl, Dagmar Divijak, Dmitrij Dobrovolskij, Anna Dybo, Thanasis Georgakopoulos, Eitan Grossman,

Martin Haspelmath, Vadim Kimmelman, Galina Kustova, Andrey Kutuzov, Asifa Majid, Matti Miestamo, Tatiana Nikitina, Stéphane Polis, Adam Schembri, Ksenia Shagal, Katalin Sipőcz, Svetlana Tolstaya, Martine Vanhove, Daniel Weiss, Wang Wenyin, Anna Zalizniak. We sincerely thank them all! We would also like to express our deepest gratitude to the Series Editor Fernando Zúñiga, the members of Editorial staff Kees Vaes and Esther Roth, and the anonymous reviewer, who took on the job of thorough reading and commenting on the first version of the manuscript. Any mistakes that remain are the sole responsibility of the authors.

References

Ameka, F. K. & Levinson, S. C. (eds). 2007. The typology and semantics of locative predication: Posturals, positionals and other beasts [Special Issue]. *Linguistics* 45(5). https://doi.org/10.1515/LING.2007.025

Apresjan, Ju. 1974. *Leksičeskaja semantika* [Lexical Semantics]. Moskva: Nauka.

Apresjan, Ju. 2000. *Systematic Lexicography*. Oxford: Oxford University Press.

van der Auwera, J. & Plungian, V. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124. https://doi.org/10.1515/lity.1998.2.1.79

Baroni, M., Bernardi, R. & Zamparelli, R. 2014. Frege in Space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* 9: 5–110.

Berlin, B. & Kay, P. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley, Los Angeles: University of California Press.

Bierwisch, M. 1967. Some semantic universals of German adjectival. *Foundations of Language* 3(1): 1–36.

Bierwisch, M. & Lang, E. 1989. *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*. Berlin: Springer. https://doi.org/10.1007/978-3-642-74351-1

Britsyn, V. M., Rakhilina, E. V., Reznikova, T. I. & Javorska, G. M. (eds). 2009. *Koncept BOL' v tipologičeskom osveščenii* [The Concept of PAIN in a Typological Perspective]. Kiev: Dmitri Burago's Publishing House.

Byrne, A. & Hilbert, D. (eds). 1997. *Readings on Color*. 2 vols. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5787.001.0001

Comrie, B. 1976. *Aspect: an Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.

Comrie, B. 1985. *Tense*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165815

Corbett, G. 1991. *Gender*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, G. 2000. *Number*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164344

Cysouw, M. & Wälchli, B. (eds). 2007. Parallel Texts. Using Translational Equivalents in Linguistic Typology. Theme Issue. *Sprachtypologie & Universalienforschung (STUF)* 60(2).

Dahl, Ö. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.

Dingemanse, M. & Majid, A. 2012. The semantic structure of sensory vocabulary in an African language. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, N. Miyake, D. Peebles & R. P. Cooper (eds), 300–305. Austin, TX: Cognitive Science Society.

Driagina-Hasko, V. & Perelmutter, R. (eds). 2010. *New Approaches to Slavic Verbs of Motion* [Studies in Language Companion Series 115]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115

Evans, N. & Wilkins, D. P. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76: 546–592. https://doi.org/10.2307/417135

Fillmore, Ch. J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, vol. 280: 20–32. https://doi.org/10.1111/j.1749-6632.1976.tb25467.x

François, A. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From Polysemy to Semantic Change*, M. Vanhove (ed), 163–215. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.106.09fra

Gage, J. 1999. *Colour and Meaning*. London: Thames and Hudson.

Georgakopoulos, Th. & Polis, S. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2): 1–33. https://doi.org/10.1111/lnc3.12270

Goddard, C. 2012. Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics* 50(3): 711–743. https://doi.org/10.1515/ling-2012-0022

Goddard, C. & Wierzbicka, A. 2007. NSM analyses of the semantics of physical qualities: *sweet, hot, hard, heavy, rough, sharp* in crosslinguistic perspective. *Studies in Language* 31(4): 765–800. https://doi.org/10.1075/sl.31.4.03god

Hardin, C. L. & Maffi, L. (eds). 1997. *Color Categories in Thought and Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511519819

Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: Oxford University Press.

Haspelmath, M. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In *The New Psychology of Language, vol. 2*, M. Tomasello (ed), 211–242. Mahwah, NJ: Lawrence Erlbaum.

Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613463

Kashkin, E. V. 2010. Semantika glagolov razdelenija ob''ekta na časti v èrzjanskom i komi-zyrjanskom jazykax (prjamye i perenosnye upotreblenija) [Semantics of verbs of cutting and breaking in Erzya and Komi-Zyryan (basic and derived meanings)]. Diploma paper, Moscow State University.

Kibrik, A. E. 1998. Does intragenetic typology make sense? In *Sprache in Raum und Zeit: Beiträge zur empirischen Sprachwissenshaft*, W. Boeder, Ch. Schroeder, K. H. Wagner & W. Wildgen (eds), 61–68. Tübingen: Narr.

Kibrik, A. E. 2009. In defence of intragenetic typology: evidence from clause coordination in Daghestanian languages. In *Eighth Biennial Conference of the Association for Linguistic Typology. Abstracts*. <http://lsa2009.berkeley.edu/alt8/kibrik_aleksandr_abstract.pdf> (26 March 2021).

Koch, P. 2008. Cognitive onomasiology and lexical change: around the eye. In *From Polysemy to Semantic Change*, M. Vanhove (ed), 107–137. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.106.07koc

Kopecka, A. & Narasimhan, B. (eds). 2012. *Events of Putting and Taking: A Crosslinguistic Perspective*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.100

Koptjevskaja-Tamm, M. (ed). 2015. *The Linguistics of Temperature*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.107

Koptjevskaja-Tamm, M., Divjak, D. & Rakhilina, E. 2010. Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In *New Approaches to Slavic Verbs of Motion*, V. Driagina-Hasko & R. Perelmutter (eds), 315–341. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115.18kop

Koptjevskaja-Tamm, M. & Rakhilina, E. V. 1999. S samymi teplymi čuvstvami (po gorjačim sledam Stokgoľmskoj èkspedicii) [With warmest regards (hot on the heels of the field trip to Stockholm)]. In *Tipologija i teorija jazyka. Ot opisanija k ob"jasneniju. K 60-letiju A. E. Kibrika* [Typology and Linguistic Theory. From Description to Explanation. For the 60th Birthday of A. E. Kibrik], E. V. Rakhilina & Ya. G. Testelets (eds), 462–487. Moscow: Jazyki russkoj kuľtury.

Koptjevskaja-Tamm, M. & Rakhilina, E. 2006. «Some like it hot»: on semantics of temperature adjectives in Russian and Swedish. *Sprachtypologie und Universalienforschung (STUF)* 59(2): 253–269. https://doi.org/10.1524/stuf.2006.59.3.253

Kruglyakova, V. A. 2010. Semantika glagolov vraščenija v tipologičeskoj perspektive [Semantics of rotation verbs in a typological perspective]. Ph.D. dissertation, Russian State University for the Humanities.

Lander, Yu., Maisak, T. & Rakhilina, E. 2012. Verbs of aquamotion: semantic domains and lexical systems. In *Motion Encoding in Language and Space*, M. Vulchanova & E. van der Zee (eds), 67–83. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199661213.003.0004

Lang, E. 2002. Spatial dimension terms. In *Language Typology and Language Universals*, M. Haspelmath & E. König (eds), 1251–1275. Berlin: de Gruyter.

Lehrer, A. 1974. *Semantic Fields and Lexical Structure*. Amsterdam: North-Holland Publishing Company.

Levinson, S. C. 2000. Yélî Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology* 10(1): 3–55. https://doi.org/10.1525/jlin.2000.10.1.3

Levinson, S. C. & Majid, A. 2014. Differential ineffability and the senses. *Mind & Language* 29: 407–427. https://doi.org/10.1111/mila.12057

Luchina, E. S., Reznikova, T. I. & Stenin, I. A. 2013. Atributivy kak istočnik grammatikalizacii: 'prjamoj' i 'rovnyj' v russkom, nemeckom i finskom jazykah [Attributives as a source for grammaticalization: 'straight' and 'even' in Russin, German and Finnish]. In *Tipología léxica*, R. Guzmán Tirado & I. Votyakova (eds), 203–209. Granada: Jizo Ediciones.

Luchina, E. S. & Nanij, L. O. 2016. Puti grammatikalizacii leksem so značeniem 'prjamoj' ['Straight' and its synonyms: paths of grammaticalization in the attributive lexeme]. *Acta Linguistica Petropolitana* 12(1): 574–587.

MacLaury, R. E. 1997. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Austin: University of Texas Press.

MacLaury, R. E., Paramei, G. V. & Dedrick, D. (eds). 2007. *Anthropology of Color: Interdisciplinary Multilevel Modeling*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.137

Maisak, T. A. & Rakhilina, E. V. (eds). 2007. *Glagoly dviženija v vode: leksičeskaja tipologija* [Verbs of AQUA-Motion: a Lexical Typology]. Moscow: Indrik.

Majid, A. 2015. Comparing lexicons cross-linguistically. In *The Oxford Handbook of the Word*, J. R. Taylor (ed), 364–379. Oxford: Oxford University Press.

Majid, A. 2021. Human Olfaction at the Intersection of Language, Culture, and Biology. *Trends in Cognitive Sciences* 25(2): 111–123. https://doi.org/10.1016/j.tics.2020.11.005

Majid, A. & Bowerman, M. (eds). 2007. Cutting and breaking events: A crosslinguistic perspective [Special Issue]. *Cognitive Linguistics* 18(2). https://doi.org/10.1515/COG.2007.005

Majid, A., Gullberg, M., Staden, M. van & Bowerman, M. 2007. How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics* 18(2): 179–194. https://doi.org/10.1515/COG.2007.007

Majid, A. & Levinson, S. C. 2007. The language of vision I: colour. In *Field Manual Volume 10*, A. Majid (ed), 22–25. Nijmegen: Max Planck Institute for Psycholinguistics.

Majid, A. & Levinson, S. C. (eds). 2011. The senses in language and culture [Special Issue]. *The Senses & Society* 6(1). https://doi.org/10.2752/174589311X12893982233551

Majid, A., Senft, G. & Levinson, S. C. 2007. The language of olfaction. In *Field Manual Volume 10*, A. Majid (ed), 36–41. Nijmegen: Max Planck Institute for Psycholinguistics.

Newman, J. (ed). 1998. *The Linguistics of Giving* [Studies in Typological Linguistics 36]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.36

Newman, J. (ed). 2002. *The Linguistics of Sitting, Standing, and Lying* [Studies in Typological Linguistics 51]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.51

Newman, J. (ed). 2009. *The Linguistics of Eating and Drinking* [Studies in Typological Linguistics 84]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.84

Partee, B. 1995. Lexical semantics and compositionality. In *An Invitation to Cognitive Science* (Second Edition). Volume 1: Language, L. Gleitman & M. Liberman (eds), 311–360. Cambridge: MIT Press.

Peirsman, Y. & Geeraerts, D. (2006). Metonymy as a prototypical category. *Cognitive Linguistics* 17(3): 269–316. https://doi.org/10.1515/COG.2006.007

Radden, G. & Kövecses, Z. (1999). Towards a theory of metonymy. In *Metonymy in Language and Thought* [Human Cognitive Processing 4], K.-U. Panther & G. Radden (eds), 17–59. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.4.03rad

Rakhilina, E. V. 2000. *Kognitivnyj analiz predmetnyx imen: semantika i sočetaemost'* [Cognitive Analysis of Concrete Nouns: Semantics and Combinability]. Moscow: Russkie slovari.

Rakhilina, E. V. 2007. Tipy metaforičeskix upotreblenij glagolov plavanija [Types of metaphorical uses of aqua-motion verbs]. In *Glagoly dviženija v vode: leksičeskaja tipologija* [Verbs of AQUA-Motion: a Lexical Typology], T. A. Maisak & E. V. Rakhilina (eds), 76–105. Moscow: Indrik.

Rakhilina, E. 2010. Verbs of rotation in Russian and Polish. In *New Approaches to Slavic Verbs of Motion*, V. Driagina-Hasko & R. Perelmutter (eds), 291–316. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115.17rak

Rakhilina, E. V. & Lemmens, M. 2003. Rusistika i tipologiya: leksičeskaja semantika glagolov so značeniem 'sidet'' v russkom i niderlandskom [Russian studies and typology: lexical semantics of verbs meaning 'to sit' in Russian and Dutch]. *Russian Linguistics* 27(3): 313–328. https://doi.org/10.1023/A:1027325127170

Rakhilina, E. V. & Reznikova, T. I. 2016. A Frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–129. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110377675-004

Rakhilina, E. V., Reznikova, T. I., Ryzhova, D. A. (eds). 2020. Tipologija glagolov padenija [Typology of verbs of falling] [Special issue]. *Acta Linguistica Petropolitana* 16(1).

Reznikova, T., Rakhilina, E. & Bonch-Osmolovskaya, A. 2012. Towards a typology of pain predicates. *Linguistics* 50(3): 421–465. https://doi.org/10.1515/ling-2012-0015

Ryzhova, D., Rakhilina, E. & Kholkina, L. 2019. Approaching perceptual qualities: the case of *heavy*. In *Perception Metaphors*, L. J. Speed, C. O'Meara, L. San Roque & A. Majid (eds), 185–208. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.19.10ryz

Shayan, S., Ozturk, O. & Sicoli, M. A. 2011. The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society* 6(1): 96–105. https://doi.org/10.2752/174589311X12893982233911

Speed, L. J., O'Meara, C., San Roque, L. & Majid, A. (eds). 2019. *Perception Metaphors*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.19

Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620904

Talmy, L. 1983. How language structures space. In *Spatial Orientation: Theory, Research and Application*, H. L. Pick & L. P. Acredolo (eds), 225–282. New York: Plenum Press. https://doi.org/10.1007/978-1-4615-9325-6_11

Vanhove, M. 2008. Semantic associations between sensory modalities, prehension and mental perceptions: A crosslinguistic perspective. In *From Polysemy to Semantic Change*, M. Vanhove (ed), 342–370. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.106.17van

Viberg, Å. 1984. The verbs of perception: a typological study. In *Explanations for Language Universals*, B. Butterworth, B. Comrie & Ö. Dahl (eds), 123–162. Berlin: Mouton. https://doi.org/10.1515/9783110868555.123

Viberg, Å. 1999. The polysemous cognates Swedish *gå* and English *go*. Universal and language-specific characteristics. *Languages in Contrast* 2: 89–115. https://doi.org/10.1075/lic.2.1.06vib

Viberg, Å. 2008. Riding, driving and traveling: Swedish verbs describing motion in a vehicle in crosslinguistic perspective. In *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*, J. Nivre, M. Dahllöf & B. Megyesi (eds), 173–201. Uppsala: Acta Universitatis Upsaliensis.

Viberg, Å. 2013. Posture verbs: A multilingual contrastive study. *Languages in Contrast* 13(2): 139–169. https://doi.org/10.1075/lic.13.2.02vib

Waag, A. 1901. *Bedeutungsentwicklung unseres Wortschatzes. Auf Grund von Hermann Pauls "Deutschem Wörterbuch" in den Hauptscheinungen dargestellt*. Lahr i.B.: M. Schauenburg.

Wälchli, B. & Cysouw, M. 2012. Lexical Typology through Similarity Semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3): 671–710. https://doi.org/10.1515/ling-2012-0021

Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

Wierzbicka, A. 1988. *The Semantics of Grammar*. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.18

Wierzbicka, A. 1999. *Emotions across Languages and Cultures*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511521256

Wierzbicka, A. 2006. Shape in Grammar revisited. *Studies in Language* 30(1): 115–177. https://doi.org/10.1075/sl.30.1.05wie

Wierzbicka, A. 2007. Bodies and their parts: an NSM approach to semantic typology. *Language Sciences* 29: 14–65. https://doi.org/10.1016/j.langsci.2006.07.002

Wilkins, D. P. 1996. Natural tendencies of semantic change and the search for cognates. In *The Comparative Method Reviewed. Regularity and Irregularity in Language Change*, M. Durie & M. Ross (eds), 264–304. Oxford: Oxford University Press.

Zalizniak, Anna A. 2008. A Catalogue of Semantic Shifts: towards a Typology of Semantic Derivation. In *From Polysemy to Semantic Change*, M. Vanhove (ed), 217–232. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.106.10zal

Zalizniak, Anna A., Bulakh, M., Ganenkov, D., Gruntov, I., Maisak, T. & Russo, M. 2012. The Catalogue of Semantic Shifts as a Database for Semantic Typology. *Linguistics*, 50(3): 633–669. https://doi.org/10.1515/ling-2012-0020

Methodology at work

Semantic fields SHARP and BLUNT

Maria Kyuseva[1], Elena Parina[2,3] and Daria Ryzhova[4]

[1]University of Surrey / [2]Institute of Linguistics, Russian Academy of Sciences / [3]Philipps-Universität Marburg / [4]HSE University, Moscow

The chapter illustrates the frame-based methodology of lexical typological analysis through the comparison of the qualities SHARP and BLUNT in 21 languages. We show that these qualities tend to be asymmetrical, with bluntness being negatively defined through sharpness. The two main oppositions found in the field are (1) the shape of a sharp object, and (2) the sense through which the quality is primarily experienced. The first opposition divides all objects into bladed (knives, etc.) and pointed (needles, etc.) ones; the second opposition contrasts touch with vision, and it further translates to the juxtaposition of function (*sharp/blunt instruments*, etc.) vs. shape (*pointed/rounded features*, etc.). We also find that these oppositions determine the semantic shifts developed by words denoting sharpness or bluntness and that the metaphoric patterns are consistent across languages.

Keywords: lexical typology, frame-based approach, semantic maps, metaphors, physical qualities, sharp, blunt

Introduction

The project of the Moscow Lexical Typology Group (Rakhilina & Reznikova 2016) on cross-linguistic comparison of qualitative features started with the analysis of two antonymous semantic fields: SHARP and BLUNT. Being very compact, these zones allowed us to elaborate a methodology for typological research of both direct and indirect meanings, and to test all our new methods and ideas in the course of their development. We will use these data to illustrate our methodology, presented from the theoretical point of view in the Introductory chapter of the present volume.

   This chapter will discuss successively all the steps of lexical typological analysis worked out by the Moscow Lexical Typology Group: defining the boundaries of

the field and constructing a questionnaire (Section 1), revealing the frame structure (Section 2), constructing a semantic map (Section 3), and investigating metaphorical extensions (Section 4). Since the fields SHARP and BLUNT appear to be asymmetrical – the former being much more lexically elaborated in the majority of languages of our sample than the latter, – we will illustrate the steps of research with the field SHARP, and we will provide a comparative analysis of its antonym BLUNT in Section 5. Section 6 presents the discussion and the comparison of the present analysis with previous studies of these physical qualities in different languages.

1.   Defining the boundaries of the field and constructing a questionnaire

Lexical typological research in our framework starts from delimitation of the field of interest: for every language of the sample, it should be made clear which lexemes are to be analysed and which ones should be excluded from the research. To achieve this, we form a list of extra-linguistic situations (presented by diagnostic contexts) within which the target lexemes could presumably occur; we use this list as a typological questionnaire that becomes the basis for the comparison of languages.

1.1   The starting point: Russian data

Our approach is data-driven in the sense that we form the list of typical situations on basis of the linguistic behavior of chosen lexemes (in contrast to the psycholinguistic approach to lexical typology; see, for example, Levinson & Wilkins 2006; Majid & Levinson 2011, and the survey in Koptjevskaja-Tamm et al. 2016). Usually, the questionnaire is constructed manually (consider, however, Ryzhova & Paperno, this volume) based on the data from three to five languages; we always start from Russian since it is the native language for all of us, which allows us to resort to our language intuition in addition to corpora and dictionary data. We choose one or several Russian lexemes and analyse their distribution in the Russian National Corpus[1] (RNC) clarifying dubious cases with the help of judgements elicited from native speakers.

In this particular case, the starting point is the word *ostryj* 'sharp'. The different occurrences of this adjective attested in the RNC correspond to various typical situations of its usage. Consider the following examples:

---

1.   https://ruscorpora.ru/new/

(1) RUSSIAN
Dlja togo,   čtoby ubedit'sja,   čto gus'   gotov,
for that:GEN.N.SG so   make.sure:INF that goose:NOM.SG ready:M.SG
*protknite   **ostrym   nožom   bedro.***
pierce:IMP.2PL sharp:INS.M.SG knife:INS.SG thigh:ACC.SG
'To ensure that the goose is cooked, pierce its leg [in the context of culinary instructions] with a **sharp knife**.'
          [Recepty nacional'nyx kuxon': Skandinavskaja kuxnya (2000–2005)]

(2) RUSSIAN
No posle togo   kak pacient   načal   žalovat'sja   na
but after that:GEN.N.SG as   patient:NOM.SG start:PST.M.SG complain:INF on
***ostruju   bol'**   v želudke,   i   ego   načalo*
sharp:ACC.F.SG pain:ACC.SG in stomach:LOC.SG and he:ACC start:PST.N.SG
tošnit',   vrači   pereveli   ego   v xirurgičeskoe
vomit:INF doctor:NOM.PL transfer:PST.PL he:ACC in surgical:ACC.N.SG
otdelenie.
department:ACC.SG
'But when the patient started complaining of **acute pain** in the stomach and vomiting, the doctors transferred him to the surgery department.'
          [Železnyj želudok (2004)// "Amurskij meridian" (Xabarovsk), 2004.12.22]

In the first example, the word *ostryj* 'sharp' describes the physical quality of being 'well sharpened, well functioning' about an instrument. The second one considers an abstract notion ('pain') instead of a physical object, and the adjective *ostryj* 'sharp' means that the experienced pain is intensive. Besides these two types of situations, this Russian adjective is used in descriptions of objects of a pointed shape (such as a nose or a shoe toe), as well as of hot, spicy meals, strong negative emotions, etc. This set of situations forms the basis of the prospective questionnaire.

Despite the fact that the situations themselves are extralinguistic, the majority of them cannot be unambiguously illustrated by means of any extralinguistic stimulus (such as a picture or a video clip). For instance, it is problematic to think of a physical stimulus that could viably represent acute pain, acute mind, or keen desire. Thus, we form a questionnaire consisting of diagnostic contexts with blanks that can be potentially filled with a lexeme belonging to the semantic field in question. Every situation is represented with several examples since we do not know in advance what level of granularity in their classification would be sufficient to capture the differences in the distribution of lexemes. While compiling the questionnaire, we aim at presenting various subtypes of every situation. For example, we illustrate the situation 'well sharpened, well functioning (about an instrument)' with the instances like 'sharp cutting instrument (knife, blade, etc.)', 'sharp piercing instrument (arrow, spear)', 'sharp natural object (stick, stone)', 'sharp body part (beak, nail)', etc.

Every illustration, i.e. every row of the questionnaire, is assigned a unique metalinguistic label of the type "qualitative feature + name of the object" since different situations usually involve different objects. However, when we work with native speakers of any language, we present them with a broader context in order to avoid ambiguity: we ask them to fill the gaps in diagnostic sentences translated into their languages. Table 1 presents a fragment of a preliminary questionnaire for the semantic field SHARP filled with the Russian data.

Table 1. Preliminary questionnaire for the field SHARP
filled with Russian data: A fragment

Situation	Context	Russian qualitative word
'well sharpened, well functioning (about an instrument)'	'sharp knife'	*ostryj*
	'sharp sword'	*ostryj*
	'sharp spear'	*ostryj*
	'sharp arrow'	*ostryj*
'having a pointed shape'	'pointed nose'	*ostryj*
'hot, spicy'	'hot taste'	*ostryj*
'strong, intensive'	'sharp pain'	*ostryj*
	…	

1.2   Intragenetic typology

Having constructed a preliminary version of the questionnaire, we turn to the data from another language. At this stage, we do not impose any restrictions on the choice of the next language except its relative accessibility. In other words, we need a language with a literary tradition, available textual corpora, and a considerable number of native speakers, since for the second language of the research we repeat almost the same procedure as for the first one. We fill the form with the data obtained from the new language and analyse the distributive properties of the words that serve as translational equivalents of the Russian lexemes. Contrary to the widely accepted view that a language sample for a typological study should be as genetically diverse as possible, we do not forbid including even closely related languages – even at the stage of questionnaire construction. This decision stems from previous evidence that lexical systems change significantly faster than grammatical ones, and hence even cognates may display different distributional properties (cf. Koptjevskaja-Tamm et al. 2010; Majid et al. 2007; Majid et al. 2015; Rakhilina 2010, and also Chapters 6 and 8 in this volume).

In the case of the field SHARP, we continue the analysis with the Serbian cognate of the Russian adjective *ostryj*, i.e. *oštar*. We find that, having much in common, these adjectives also demonstrate a number of different usage patterns. For example,

the Serbian *oštar* can describe lines, pictures, and contrast (in these contexts it means 'clear, precise, sharp', Example (3)), or beards, cactuses and bushes ('prickly', (4)), while Russian *ostryj* does not occur in the corresponding contexts. On the contrary, Russian *ostryj* can describe the taste, while Serbian uses another lexeme in this case, i.e. *ljut*. In its literal use, the adjective *ljut* means 'evil, wicked, angry (about a human being)', and the meaning 'hot, spicy' presents a metaphorical extension of this lexeme.

(3) SERBIAN
    Kad   se   objektiv   fotoaparata   podesi   na
    when REFL lens:NOM.SG camera:GEN.SG adjust:PRS.3SG on
    beskonačnu   udaljenost,   na površini   filma   se
    infinite:ACC.F.SG distance:ACC.SG on surface:LOC.SG film:GEN.SG REFL
    *pojavljuje   **oštra**   **slika**   onih*
    appear:PRS.3SG sharp:NOM.F.SG picture:NOM.SG that:GEN.M.PL
    detalja   koji   su   na jako velikoj
    detail:GEN.PL which:NOM.M.PL be:PRS.3PL on very big:LOC.F.SG
    udaljenosti.
    distance:LOC.SG
    'If you push the camera lens at the maximum length, you'll see **sharp outlines** of the distant objects in the picture.'

(4) SERBIAN
    Dlaka   *je*   *gusta,*   **oštra,**   *tvrda*
    hair:NOM.SG be:PRS.3SG thick:NOM.F.SG sharp:NOM.F.SG hard:NOM.F.SG
    poput čekinja,   donja   dlaka mekana   i   odlično
    like   bristles:GEN.PL lower:NOM.F.SG hair   soft:NOM.F.SG and perfectly
    ga   štiti   od   nevremena.
    he:GEN protect:PRS.3SG from bad.weather:GEN.SG
    'Its guard **hair** is thick, **prickly** and as hard as bristle, while its undercoat is soft and keeps it warm in bad weather (about a dog).'

1.3   The "Shuttle" method and the core contexts

In the next step, we extend the questionnaire with the new contexts revealed through the Serbian data and then return to Russian to fill in the empty slots. Table 2 shows a fragment of the updated questionnaire filled with the Serbian and Russian data. We use the term "shuttle method" to refer to this helix-like design of the research: every new language brings in some new evidence that needs to be analysed in the previously studied languages. We believe that this method safeguards us from the danger of excessive influence of the first language: the "shuttle method" ensures that we should end up with the same questionnaire regardless of the language we

Table 2. Updated questionnaire for the field SHARP

Context	Russian	Serbian
'sharp knife'	*ostryj*	*oštar*
'sharp sword'	*ostryj*	*oštar*
'sharp spear'	*ostryj*	*oštar*
'sharp arrow'	*ostryj*	*oštar*
'pointed nose'	*ostryj*	*oštar*
'spicy taste'	*ostryj*	*ljut*
'sharp pain'	*ostryj*	*oštar*
'sharp line'	*rezkij, četkij*	*oštar*
'prickly bush'	*kolyučij*	*oštar*
…		

started from; see a similar approach in (François 2008). The process of adding new contexts to the questionnaire is not infinite. Some languages have idiosyncratic metaphorical extensions, but it would be of little use to include them all in the questionnaire. Our experience shows that a set of typologically recurrent meanings can be formed after a fine-grained examination of three to five languages (even if they are related to each other).

The contexts in the questionnaire have unequal statuses in a given semantic field. We divide them into the core and the peripheral group. The core contexts represent the central meanings of the field which we define according to the following principles:

1.   these meanings correspond to easily definable physical situations;
2.   among physical situations, we choose those which tend to colexify
      with each other.

The peripheral contexts illustrate abstract ideas and physical situations that are rarely covered by lexemes from the field.

In the semantic field SHARP, the core meanings are 'well-functioning' (about (quasi-)instruments) and 'pointed shape'. The first meaning is illustrated by such contexts as *sharp knife, sharp arrow, sharp stick*; the second meaning is realized in the phrases *pointed cap, pointed nose, pointed top of the mountain*, among others. Our analysis is different from some other analyses of this semantic field (see an overview in Section 6): we do not include the meanings 'prickly' and 'hot, spicy (taste)' in the set of the core meanings despite the fact that they represent easily definable physical situations. Our reasoning behind this decision is the following: typologically, these meanings are rarely covered by lexemes from the field SHARP, i.e. they are rarely colexified with the meanings 'well-functioning' (about

(quasi)-instruments) and 'pointed shape'. Instead, they are often expressed with lexemes coming from different semantic fields, such as EMOTIONS (*ljut* 'angry' in Serbian to denote spicy food), TEMPERATURE (*hot* for taste in English), and others.

After establishing the boundary between the core and the peripheral contexts on the data of three to five languages, we use it to select lexemes for the analysis in every new language. Only lexemes that can occur in one or more core contexts in their literal meaning are labelled as belonging to the field and, therefore, get analysed. For example, in Russian, the only lexeme in the field SHARP is *ostryj*. The words *koljučij, rezkij, četkij*, although attested in the questionnaire (see Table 2), are not selected into the sample, because they do not express any of the core meanings of this semantic field.

2.   Revealing the frame structure

Having designed the questionnaire, we expand the sample by adding data of new languages. We consult dictionaries and text corpora (in case they are available) and carry out elicitation sessions with native speakers. Since our questionnaire is a list of contexts, we translate the diagnostic noun phrases into the target language; we usually use broader contexts which unambiguously illustrate the underlying situations for the elicitation task (such as 'I cannot see anything behind the high wall' for the context 'high wall').

As a result, we obtain a table with contexts as rows and lexemes as columns: for every lexeme, we indicate whether it can be used in a given context. For the field SHARP, the table consists of 136 rows, see a fragment below in Table 3:

Table 3. Filled lexical typological questionnaire for the field SHARP: A fragment

	Russian	Chinese	Hungarian	French		Serbian
	ostryj	*jianrui*	*szuros*	*pointu*	*tranchant*	*oštar*
'sharp knife'	1	0	0	0	1	1
'sharp sword'	1	0	0	0	1	1
'sharp sabre'	1	0	0	0	1	1
'sharp blade'	1	0	0	0	1	1
'sharp needle'	1	1	0	1	0	1
'sharp arrow'	1	1	0	1	0	1
'sharp spear'	1	1	0	1	0	1
'pointed nose'	1	0	0	1	0	1
'pointed elbow'	1	0	0	1	0	1
'pointed beak'	1	0	0	1	0	1
'pointed peak'	1	0	0	1	0	1

The language sample consists of 21 languages:

1–2.   Slavic: Russian, Serbian;
3–4.   Romance: French, Italian;
5–6.   Germanic: English, German;
7.   Celtic: Welsh;
8–11.   Uralic: Finnish, Hungarian, Izhma Komi, Moksha;
12.   Indo-Iranian: Hindi;
13–14.   Caucasian: Aghul, Besleney Kabardian;
15.   Japonic: Japanese;
16.   Koreanic: Korean;
17.   Sino-Tibetan: Chinese;
18.   Austronesian: Malay;
19.   Mande: Kla-dan;
20.   Isolate: Basque;
21.   Russian Sign Language.

Our findings reveal that some rows of the table filled with language data are very close to each other in the sense that they have either identical or very similar content. For example, if a lexeme can be used in the context of 'needle', it can also describe an arrow; if, on the contrary, a lexeme does not co-occur with 'needle', the combination with 'arrow' is also forbidden (see Table 3). These clusters of rows with similar content form "frames", i.e. typical situations described by lexemes of the field. In our theory, frames are basic organisational blocks of semantic fields that serve as the basis of the cross-linguistic comparison. See a detailed discussion of this notion in Chapter 1 of this volume.

The core contexts of the field SHARP cluster in three core frames: 'sharp (about objects with a cutting edge; rows 1–4 of Table 4)', 'sharp (about objects with a piercing point; rows 5–7)', 'sharp, pointed shape' (rows 8–11).

These clusters form a "conceptual system" based on two oppositions: 'line' versus 'point' and 'function' versus 'shape'. The first opposition contrasts sharp objects with a cutting edge to sharp objects with a piercing point & objects of a pointed shape. While in the former ones, the line projection of the blade is profiled, the latter ones have a highlighted point at the end (i.e., end of the needle, top of the mountain). The second opposition contrasts objects of a pointed shape to the two remaining frames. When a lexeme from the field SHARP is used with names of objects of pointed shape, it describes the form of the object. When a lexeme from this field is used with names of instruments, either cutting or piercing, it describes the quality of the object's functioning.

Table 4. Core frames in the semantic field SHARP

		Russian	Serbian	Chinese		French	
		ostryj	*oštar*	*jiānruì*	*jiāntóu*	*pointu*	*tranchant*
1	'sharp knife'	1	1	0	0	0	1
2	'sharp sword'	1	1	0	0	0	1
3	'sharp sabre'	1	1	0	0	0	1
4	'sharp blade'	1	1	0	0	0	1
5	'sharp needle'	1	1	1	0	1	0
6	'sharp arrow'	1	1	1	0	1	0
7	'sharp spear'	1	1	1	0	1	0
8	'pointed nose'	1	1	0	1	1	0
9	'pointed elbow'	1	1	0	1	1	0
10	'pointed beak'	1	1	0	1	1	0
11	'pointed peak'	1	1	0	1	1	0

As for peripheral meanings, they are less structured cross-linguistically but also form some typologically relevant patterns. For example, the contexts 'sharp vision', 'good hearing', and 'keen scent' illustrate a single frame that can be called 'well-functioning (about perceptional physiological systems)'. For the lexemes of the field SHARP, this meaning is figurative since, in such cases, a qualitative word 'sharp' contributes only an abstract evaluative meaning to an utterance and does not presuppose either physical impact[2] or a specific shape. We regularly attest this metaphorical extension in genetically and areally diverse languages: compare Kabardian *ne žan* 'sharp vision' (lit. "sharp eye"), Japanese *shita surudoi* 'keen scent' (lit. "sharp scent"), and Komi *jues s'in* 'sharp vision' (lit. "sharp eye").

---

2. Of course, there is some physical impact in situations of seeing, hearing, or smelling. For example, when we see an object, we physically register its presence with our eyes. However, this is a fundamentally different type of impact from a cutting knife or a piercing spear since it does not presuppose an action that would physically affect the referent.

Technically speaking, frames are minimal meanings that are distinguished lexically in at least one language of the sample. In the field SHARP, as Table 4 shows, every frame covers a group of contexts with identical rows, i.e. with the same sequences of ones and zeros. The resulting "minimal meanings" are data-driven: different data would lead to different framing. For example, if the French word *tranchant* did not describe the situation 'sharp knife' (if the cell in the first row & last column of the table was 0 and not 1), this would lead to the splitting of the frame 'sharp instruments with a cutting edge' into two parts. Potentially, any new language can trigger this sort of splitting. In practice, however, we have not found this to happen: frames acquire a distinct shape after a thorough analysis of three to five languages. The reason behind this is the semantic motivation: frames are not formed randomly but combine contexts with the same situation types, which reflect the human experience.

In the following section, we discuss different strategies in the lexicalization of the core frames, and Section 4 examines cross-linguistic regularities within the peripheral meanings of the field in more detail.

3.   Constructing a semantic map

We consider revealed frames of the field to be a part of the universal inventory of lexical meanings that every language clusters in its own way (similarly to the universal grammatical inventory, Bybee & Dahl 1989). The number of possible clusterings is strictly limited, and to demonstrate all the regularities and constraints we use semantic maps – a tool which has been widely used in grammatical typology (cf. Lazard 1981; Haspelmath 1997, 2003; van der Auwera & Plungian 1998; Cysouw et al. 2010), and further extended to the cross-linguistic analysis of lexical domains (François 2008; Georgakopoulos et al. 2016; Maisak & Rakhilina 2007, see also Georgakopoulos & Polis 2018 for a recent in-depth overview).

A semantic map is a graph featuring grammatical or lexical meanings as nodes, which are linked with lines, or "edges", if these nodes can be covered by the same linguistic item in at least one language of the sample. The contiguity of nodes (in our case, frames) is significant. Closeness reflects possibilities for colexifications: the closer the meanings are on the map, the more likely they are to be covered by the same lexical means in some language. The connectivity restriction imposed on semantic maps in grammatical typology (cf. Croft 2001: 96) is also valid in our approach to the analysis of lexicon: any lexeme should denote only a connected subset of possible meanings.

Figure 1 represents the semantic map for the core meanings of the field SHARP. It demonstrates that certain combinations of frames are ruled out: we hypothesize that there are no lexemes that would simultaneously describe sharp instruments

with a functional edge (knife, saw) and objects of a pointed shape (nose, cap, toe) while leaving aside sharp instruments with a functional end-point (arrow, spear).
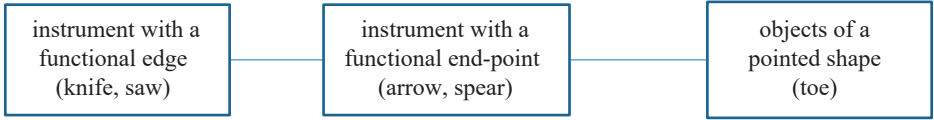
instrument with a functional edge (knife, saw)	instrument with a functional end-point (arrow, spear)	objects of a pointed shape (toe)

Figure 1. Semantic map of the field 'sharp'

3.1   Types of the SHARP systems

SHARP frames are lexicalized according to four main strategies: dominant, complex, binary, and reduced (in terms of Maisak & Rakhilina 2007). In languages that follow the dominant strategy, there is only one main lexeme covering the field. We call it the "dominant" lexeme of the field. This pattern is represented, among others, in English, Russian, Serbian, Finnish, and Malay (Figure 2).

instrument with a functional edge (knife, saw)	instrument with a functional end-point (arrow, spear)	objects of a pointed shape (toe)	Russian: *ostryj* English: *sharp* Serbian: *oštar* Finnish: *terävä* Malay: *tajam*
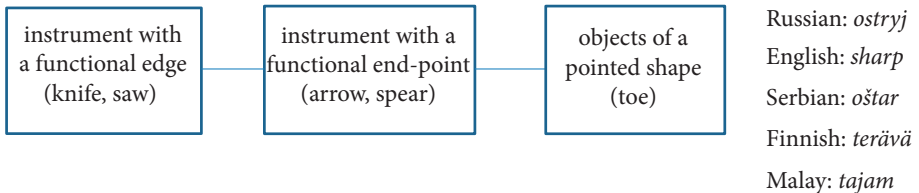
Figure 2. Dominant system of the field SHARP

A complex system is found, for example, in Mandarin Chinese. In this language, the field is covered by seven lexemes (Figure 3).

instrument with a functional edge (knife, saw)	instrument with a functional end-point (arrow, spear)	objects of a pointed shape (toe)

jianrui, ruili, jianli, fengli   *jianxing, jiantou*

jian
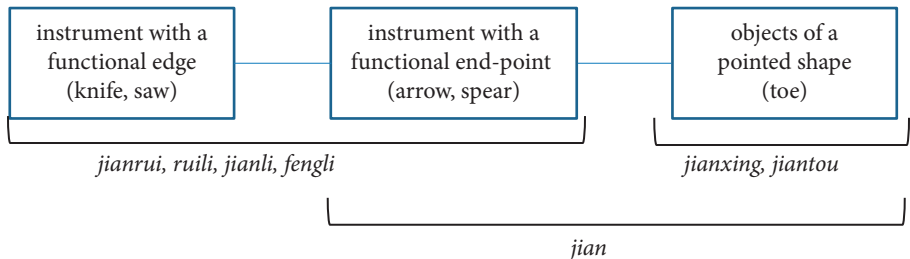
Figure 3. Complex system of the field SHARP: Mandarin Chinese

The most commonly found strategy in our data for the field SHARP is the binary one, i.e. a system with two lexemes covering the field. Logically, there are two possible ways to divide three linearly organised frames into two clusters; both of them are represented in natural languages and correspond to the underlying oppositions

we discussed earlier. In the first subtype of binary systems, one lexeme describes well-functioning instruments, while another describes objects of a pointed shape. This is the expression of the opposition 'function' versus 'shape'. Japanese and Welsh, for example, employ this strategy.

instrument with a functional edge (knife, saw)	instrument with a functional end-point (arrow, spear)	objects of a pointed shape (toe)

Japanese: *surudoi*   *togatta*
Welsh: *miniog*   *pigog*
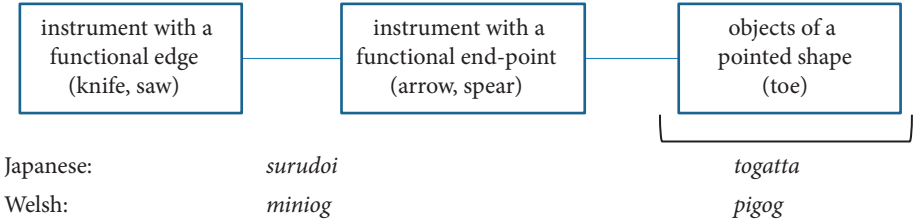
Figure 4. Binary system based on the opposition 'function' vs. 'shape'

In the other subtype, the first lexeme describes cutting instruments, while the second one covers piercing instruments and objects with a salient point. This strategy reflects the opposition 'line' versus 'point'. It is used in German, Hungarian, Komi, Kabardian, and Kla-Dan, among others.

instrument with a functional edge (knife, saw)	instrument with a functional end-point (arrow, spear)	objects of a pointed shape (toe)

German *scharf*   *spitz*

Komi: *lečyd*   *jues*

Kla-dan: líẽ̋ẽe   *zű̃ẽ̈ẽ*
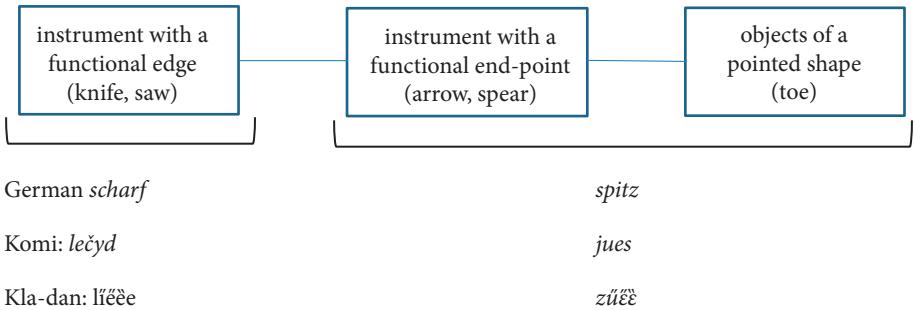
Figure 5. Binary system based on the opposition 'line' vs. 'point'

We predict that if a language has two words covering the field SHARP, one of them will cover instruments with a functional edge (the leftmost frame on the scheme), and the other one – objects of a pointed shape (the rightmost frame). The clustering of the middle frame may vary: it can potentially be covered by either lexeme.

In reduced systems, one or two core frames lack a dedicated lexical means. Within our sample, two languages manifest this strategy: Basque and Aghul. Both of them feature one lexeme with the meaning 'sharp' that includes the frame 'sharp cutting instrument'. In Basque, this lexeme also describes instruments with a piercing point, but there is a lexical gap on the 'pointed shape' meaning. In Aghul, neither 'piercing point' nor 'pointed shape' can be denoted by a one-word expression.
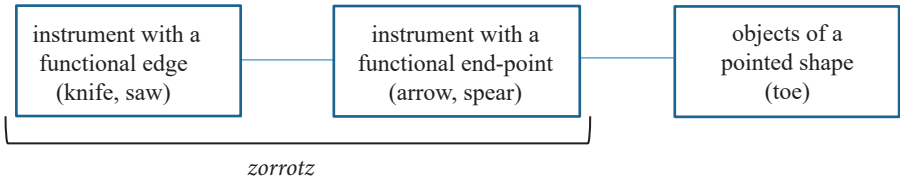
Figure 6. Semantic field SHARP in Basque
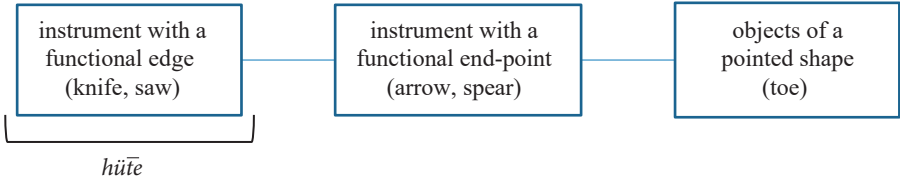
Figure 7. Semantic field SHARP in Aghul

3.2   Intermediate cases

A semantic map is a scheme that, to a certain extent, simplifies reality. Our data show that frames are not discrete: they form a continuous space with certain focal points. These points serve as the nodes of a map, but there are also some intermediate cases. For example, the phrase *sharp claws* can be translated into Italian either with the lexeme *affilato*, which is normally used with lexemes for cutting instruments, or with the lexeme *appuntito*, which is combined with words for piercing instruments. This can be explained by different conceptualizations of the object 'claw': on the one hand, claws narrow towards their ends, which likens them to instruments with a sharp point; but, on the other hand, they produce scratches rather than pointed wounds, which brings them closer to instruments with a sharp edge; this is illustrated in Figure 8.
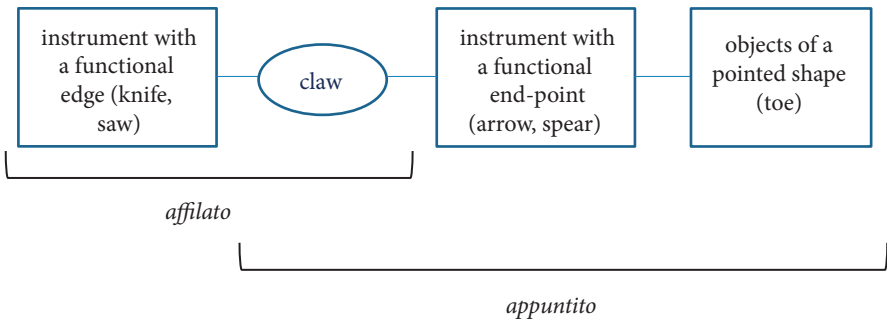
Figure 8. An intermediate case in the field SHARP: Italian

Another example of an intermediate case is the meaning 'sharp elbow', which can either behave as an instrument (cf. *he hit me with his sharp elbow*), in which case it is described by a functional lexeme; or it can be conceptualized as an object of a specific shape, in which case it is described by a lexeme that covers the frame 'objects of a pointed shape'. This pattern is observed, for example, in Japanese; see Figure 9.
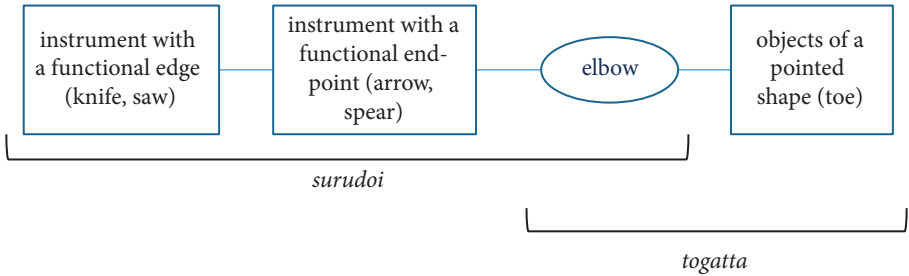
Figure 9. An intermediate case in the field SHARP: Japanese

4.   Metaphorical extensions

As Rakhilina & Reznikova (2016, see also this volume) argue, our methodology offers the possibility of analysing not only literal meanings expressed by the words in question but also their semantic extensions. By using the typological question-naires described in Section 1, we collect data about both the literal and the figurative usages of lexemes. All the semantic fields studied so far demonstrate striking recurrence of polysemy patterns in different languages. This holds for the field SHARP as well. Although some words belonging to this field incidentally have idiosyncratic metaphors (such as 'fidgety' (about a child) in Komi, and 'narrowly focused' (about a study) in French), the majority of semantic shifts are replicated from language to language. Overall, we found 25 recurrent patterns.[3]

Moreover, these figurative meanings can typically be predicted from the taxonomical class of the governing noun, which is in line with the distributional

---

3.   The figurative uses we identified are: 1. well-functioning, quick to understand things (mind); 2. clever, witty (person); 3. correct, right to the point (words); 4. perceptive (look); 5. difficult, bad (temper); 6. harsh (person); 7. mean, rough, offending (words); 8. hostile (look); 9. strong, intense (pain); 10. affecting sensory organs (sound, light, colour, smell, taste); 11. violent, fierce (interactions: conflict, argument, fight); 12. grave, dangerous (disease); 13. quick to notice, good (eyesight, hearing, intuition); 14. urgent (problem, issue); 15. at a sharp angle (turn); 16. sudden, rapid (movement, kick, stroke); 17. sudden, rapid (abstract: change, rise, decline); 18. at a high speed (running, racing); 19. steep (hill, slope, bank); 20. prickly (bristle, jumper, blanket); 21. biting, harsh (wind); 22. biting, nipping (frost); 23. clear (line, contrast); 24. clear, well-organized (order, argument, explanation); 25. strong (emotions).

hypothesis (Sahlgren 2008). For example, if a lexeme with the meaning 'sharp' is used together with a word meaning 'pain', the resulting meaning is always 'intensive pain', irrespective of the language in which this phrase occurs. Similarly, if a 'sharp' lexeme is governed by a noun meaning 'change', the resulting meaning will always be 'sudden, abrupt change'. Sometimes, however, a noun may give rise to a wide range of possible metaphors with the lexeme in question. For example, a phrase which can be literally translated into English as 'a sharp person' means 'a strict person' in Serbian (adj. *oštar*), 'a straightforward person' in French (*tranchant)*, 'a harsh/rude person' in Italian and Basque (*tagliente* and *zorrotz*, correspondingly), 'a shrewd/sagacious person' in Finnish and Japanese (*terävä* and *surudoi*, correspondingly), and 'an active person' in Kabardian (*ž'an*). This inconsistency occurs in the zone related to human appearance and behavior: the resulting meaning is extremely hard to predict with such nouns as 'person', 'look', 'word', 'character', and some others. We believe that the reason for this is the anthropocentric nature of language (Rakhilina 2010), which generally means that "the human" is a special zone for languages. Unlike artifacts, people possess a wide range of abstract qualities: kindness, honesty, insight, sense of humor, and others. Therefore, when applied to a human, a property-concept lexeme can potentially highlight any of these qualities.

Semantic extensions in the semantic field SHARP can build on different aspects of literal usages. Let us illustrate this with the phrase 'sharp knife'. On the one hand, a sharp knife cuts well, in other words, it is good at performing its function. In a similar way, a 'sharp mind' is able to think of relevant concepts and logical relationships. On the other hand, a sharp knife can harm a person and cause a very unpleasant feeling. Similarly, 'sharp sounds' or 'sharp words' can impair our hearing or negatively affect our emotions. Finally, a sharp knife leaves a distinct thin groove on a surface. This is similar to a picture with 'sharp lines'. These situations illustrate the three main metaphorical clusters in the field SHARP, namely: functional, experiential, and visual metaphors. The functional branch embraces figurative meanings in which an abstract concept or a concrete object is conceptualized or acts as a thing that is good in performing its function. The most common examples of such meanings, besides 'clever mind' (5), are 'acute sight/sense of smell' (6) and 'witty words' (7).

(5) ENGLISH
     *Even at the age of twelve, Hertzel had a **sharp brain** and a ready answer to all emergencies.*

(6) KOMI
     Etija ponmys   jues   pel'a.
     this   dog:POSS.3SG sharp ear:ATTR
     'This dog has **acute hearing** (lit., "sharp ear").'

(7) RUSSIAN
S   nim   bylo   črezvyčajno prijatno   besedovat' – obmenivat'sja
with he:INS be:PST.N.SG extremely   pleasantly converse:INF exchange:INF
*korotkimi, **ostrymi   zamečanijami.***
brief:INS.PL sharp:INS.PL remark:INS.PL
'It was sheer pleasure to talk to him and to exchange brief witty (lit. "sharp")
comments.'

In experiential metaphors, a referent is compared to a sharp instrument as a dangerous object which can hurt a person physically or emotionally, for example, 'sharp pain' (8), 'harsh light/sound' (9), and 'urgent problem' (10).

(8) FRENCH
*Vous devez chercher l'attention médicale d'urgence si vous sentez la **douleur aiguë**
dans le haut de l'estomac.*
'If you feel **acute pain** in the upper abdomen call ambulance.'

(9) FINNISH
Parasta   on,   jos huoneilma   on
better:PART.SG be:PRS.3SG if   room.temperature:NOM.SG be:PRS.3SG
raikasta   ja valo   pehmeää.   Veto   tai
fresh:PART.SG and light:NOM.SG soft:PART.SG draught:NOM.SG or
terävä   valo   häiritsevät   rauhoittumista.
harsh:NOM.SG light:NOM.SG bother:PRS.3PL relaxation:PART.SG
'It is better when a room has fresh air and soft light in it. A draught or **harsh light** will negatively affect relaxation.'

(10) RUSSIAN
*Odna   iz naibolee **ostryx   problem** –*
one:NOM.F.SG of most   sharp:GEN.PL problem:GEN.PL
kadrovyj   deficit.
personnel:NOM.M.SG shortage:NOM.SG
'One of the most **urgent** (lit. "sharp") **problems** we are facing is the shortage of staff.'

Finally, visual metaphors are based on the visual perception of cuts made by a sharp instrument as being thin, clear, and easily distinguishable. We observe such metaphors as 'clear line (physical and abstract)' (11) and 'clear picture' (3) in this zone.

(11) ENGLISH
*Of course, it's sometimes hard to draw a **sharp line** between form and function.*

It can be argued that this meaning expresses the same type of cognitive impact as the core frame 'objects of a pointed shape' and, thus, must be placed onto the

semantic map of core meanings. There are, however, at least two reasons to treat this meaning differently. First, in the majority of languages from our sample, it is covered by a lexeme that does not express an idea of sharpness. Similarly to the meaning 'hot, spicy meal', this offers indirect evidence in favour of the peripheral status of this usage. Second, unlike 'objects of a pointed shape', this metaphor can be applied not only to a physical object but also to an abstract concept, see, e.g. (11).

Besides defining the range of metaphors occurring in a given semantic field, our methodology allows us to establish connections between them and the core frames. For some figurative meanings, we can even identify the particular physically-based frame they derive from. The meaning 'clear line', for example, is connected to the frame 'instrument with a cutting edge', while the meaning 'penetrating look' originates from the frame 'instrument with a functional end-point'. Such connections are never established merely on the basis of the speculative plausibility of the relationship. Rather, they are revealed by the data themselves: if a lexeme covers only one core frame in the field, then all its metaphors are (directly or indirectly) connected to this frame. For example, if a lexeme X in a language Y describes only one physical frame – 'sharp instruments with a cutting edge' – and develops a metaphor 'with clear-cut lines' (about a picture or a photo), then we can assume that this metaphor is derived from the frame 'sharp instruments with a cutting edge'. If we encounter more than one case like this in our sample, and if we do not encounter any lexemes that have this metaphor but do not cover the core frame in question (and cover either the frame 'sharp instrument with a piercing end-point' and/or the frame 'pointed shape' instead), this strengthens our assumption.

Such connections between core frames and metaphors can only be drawn from lexemes that limit their core usages by one frame. Dominant systems with one lexeme spreading across all core frames cannot be used to establish these connections. If a lexeme X in a language Y, for example, has the metaphor 'acute' (about pain) and covers all three core frames – 'sharp instruments with a cutting edge', 'sharp instruments with a penetrating end-point', and 'pointed objects' – then we do not know which frame gave rise to the metaphor. Since the majority of languages in our sample have the dominant strategy of lexicalization of the field SHARP, we are not able to establish the connections between all the observed metaphors and the core frames. A much larger sample of languages is needed for this. For some semantic extensions, however, the source meaning is clear. For example, the link between the meaning 'clear cut' (lines) and 'sharp instruments with a cutting edge' is evidenced in the English lexeme *razor-sharp*. It possesses the metaphorical extension in question (12), and it covers exclusively the core frame 'instrument with a functional cutting edge' in its literal usages. Correspondingly, the connection 'penetrating look' – 'instrument with a functional end-point' is supported by the French lexeme *aigu* (13).

(12)   ENGLISH
    *One of the hardest things in photography is to get **razor-sharp images**.*

(13)   FRENCH
    *Mais elle a un **regard aigu**; une blessure s'ouvre en moi.*
    'But she has a **keen look** in her eyes; a wound opens inside me.'

While lexemes covering either of the two functional frames ('instrument with a functional edge' and 'instrument with a functional end-point') are regularly used in metaphorical contexts, none of the metaphors in the field SHARP seems to be derived from the core frame 'objects of a pointed shape'. We can plausibly argue this because, within our sample, the lexemes that specialize in this frame do not feature any metaphorical extensions.

The abundance of dominant strategies in the field SHARP in our sample and the consequent lack of clearly visible connections between the majority of metaphors and the core frames make it impossible to include metaphorical frames in the semantic map at this stage of the research. The data at hand allow us to reveal general tendencies of the semantic shifts in this field and the conceptual motivations behind them. A much larger sample would be necessary to enable a structural representation of these meanings and, therefore, their semantic mapping.

5.   Semantic field BLUNT

Usually, we analyze words not by themselves, but in comparison with their synonyms or antonyms. This choice is motivated by the fact that complete synonyms and antonyms are very rare in languages. Much more often, a synonym/antonym shares only a part of its meanings with a target word. For example, both the English adjectives *large* and *big* can describe an object's size (14–15), but only *big* can have the meaning 'loud' (16):

ENGLISH
(14)   *After passing through an oak grove, we spied the **large cage** a hundred yards ahead.*

(15)   *One of them held a **big cage**, with an even bigger bird jammed into it.*

(16)   "*No patient before or after has made such a **big/*large noise**,*" he says in a voice filled with both awe and laughter.

A comparison of contexts in which synonyms or antonyms can occur contributes significantly to revealing their underlying semantic structure. Therefore, we have analyzed not only the field SHARP but also its antonym BLUNT in all the languages of the sample.

Perhaps, the most surprising finding from this analysis is that the field BLUNT has a less elaborate lexical coverage in comparison to its antonym. This could not have been deduced based on Russian only, which was the first language of the research project. In Russian, the two semantic fields in question are covered by the dominant adjectives *ostryj* and *tupoj*, correspondingly. Both lexemes are very frequent, and both exhibit complex structures of metaphorical extensions. The other languages of the sample, however, tend to use more limited resources for expressing the meaning 'blunt'. For example, in Mandarin Chinese, seven lexemes in the field SHARP correspond to only one lexeme in the field BLUNT. In French, SHARP is covered by three lexemes, while BLUNT is not covered by single-word lexemes at all. To express this meaning, participles and paraphrase sentences are used instead (*un couteau qui ne coupe pas,* lit. 'a knife which does not cut'; *un couteau emoussé*, lit. 'a blunted knife'). Overall, in 15 out of the 21 languages, the lexemes for describing SHARP outnumber the lexemes for BLUNT; five languages have an equal number of lexemes in both fields, and only one language of the sample, Finnish, has more words expressing the quality BLUNT than its antonym. The quality BLUNT contains three direct frames, which form the semantic map (note that these frames correspond to the three core frames of the field SHARP):
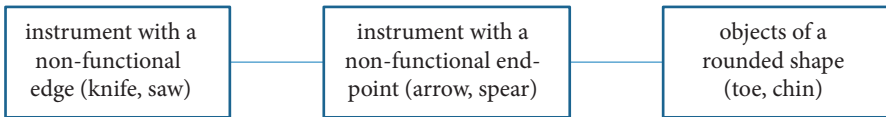
Figure 10. Semantic map of the field BLUNT

Different lexicalization strategies of BLUNT are observed in our sample. For example, Russian, Serbian, German, and Basque use one dominant lexeme to describe all the three situations: malfunctioning cutting instruments, such as a blunt knife, saw, or sword; malfunctioning piercing instruments, such as a blunt needle, awl, or spear; and objects of a rounded shape, such as a chin or a toe of the boot (i.e. the lexemes *tupoj, tup, stumpf,* and *kamuts,* respectively). Mandarin Chinese, Japanese, Korean, and Malay have one lexeme in this field as well, but it covers only two frames of the map: cutting and piercing instruments. As for the rounded shape, it is described with paraphrases or lexemes primarily covering a different field. Komi, Kabardian, and Finnish demonstrate the distributive strategy with two lexemes, the first lexeme covering the leftmost frame and the second one describing the right and the middle frames (i.e. the lexemes *nyž* vs. *tuč*; *ʒagʷe* vs. *pagʷe*; *tylsä* vs. *tylppa*, respectively). As the reader may notice, this strategy reflects the 'line' versus 'point' opposition. In Aghul and Kla-Dan, we observe a reduced system, where the only lexeme found in the field covers only one frame (*q̄üre* and *sĕsử*, respectively). Interestingly, in both

languages, the frame covered by the lexeme is 'cutting instruments'. This behavior is similar to what we find in the reduced systems of the field SHARP, and it indirectly suggests that 'cutting instruments' is the most salient frame for these two qualities. And finally, French demonstrates a lexical gap in the field of BLUNT.

The lesser degree of complexity of the field BLUNT can also be observed in the domain of figurative meanings. The list of metaphors associated with the lexemes expressing the meaning 'blunt' is much narrower than that of their antonyms. It consists of five clusters (compared to 25 in the field SHARP):

1. foolish, stupid (about a person);
2. poor, diminished sense of smell/eyesight;
3. slow, sluggish (for example, 'sluggish sales', 'sluggish traffic');
4. lasting, pressing (about pain);
5. dull sound/colour/light.

The first two metaphors from the list are observed in Russian (*tupoj*), Serbian (*tup*), Italian (*ottuso*), Finnish (*tylsä*), Japanese (*nibui*), and Basque (*kamuts*); see Examples (17)–(18). The meaning 'slow, sluggish' is expressed by the Japanese lexeme *nibui* (19) and, partially, by the Italian word *ottuso*. 'Pressing pain' is associated with the field BLUNT in Russian, Finnish, Serbian, and Japanese (20). Finally, the last metaphor of the list is observed in German (*stumpf*) in (21), Serbian (*tup*), and some other languages.

(17) BASQUE
*Erantzun zuzena   emango   luke baita **kamutsenak**   ere.*
answer   right:DET give:PFV.FUT AUX PTCL stupid:SUP.SG.ERG ADD
'Even the most **stupid person** could give the right answer.'

(18) SERBIAN
*Onda ide   lupanje   srca,   **tup***
then go:PRS.3SG throbbing:NOM.SG heart:GEN.SG blunt:NOM.M.SG
***sluh**,   glavobolja   i   ako ne reagiram   na*
hearing:NOM.SG headache:NOM.SG and if   not react:PRS.1SG on
vrijeme   zacrni   mi   se   pred očima.
time:ACC.SG blacken:FUT.3SG me:DAT REFL before eye:INS.PL
'Hence, I have increased heartbeat, headache, and a **dull hearing**, and if I do not react in time the world will go dark before my eyes.'

(19) JAPANESE
*Mausu ka:soru no **hanno:**   ga **nibu-ku** nat-ta   toki…*
mouse cursor GEN response NOM dull-CNV become-PST time
'when your mouse pointer has become slow to respond (lit. response has become dull)'

(20) FINNISH
Oikean   *kylkikaaren*   *alla*   *tuntuu*   ***tylppä***
right:GEN.SG rib.curve:GEN.SG below:POSTP feel:PRS.3SG dull:NOM.SG
kipu   *ja*   *tosi* *epämukava*   *tunne.*
pain:NOM.SG and very unpleasant:NOM.SG feeling:NOM.SG
'There is a **dull pain** in the right upper abdomen and a very unpleasant feeling.'

(21) GERMAN
Mit gebrochenen, ***stumpfen Farben,*** *zum Beispiel Oliv, wissen Kinder nichts anzufangen.*
'Children cannot do anything if they are surrounded by greyish, **dull colors** as, for example, olive-green.'

These metaphors can be divided into functional and experiential groups (cf. Section 4). Similarly to a sharp knife, a blunt knife can entail different perceptive images. On the one hand, it does not cut well; it is inefficient in performing its function. This is similar to a mind that is not keen enough (Metaphor 1) and to eyesight or sense of smell which is not sufficiently acute (Metaphor 2). The third metaphorical situation belongs to the same group of functional metaphors: sluggish trading, for example, is trading which does not go as actively as it should. On the other hand, a blunt knife can cause an unpleasant sensation. This aspect gives rise to the group of experiential metaphors. We cannot cut ourselves with a blunt knife, but if pushed against our skin for a certain amount of time it will cause a pressing pain, which motivates the fourth metaphor. The experiential group also contains the last metaphor of the list: as well as a pressing pain, a dull sound is constant and not strongly expressed.

The abundance of dominant lexicalization strategies in the field did not allow us to establish empirical connections between different metaphors and direct frames; a larger language sample would be necessary in order to do this. It is evident, however, that BLUNT is less developed than its antonym in both direct and figurative meanings. This, together with the relatively smaller number of lexemes describing this quality in languages, suggests a secondary status of this field with respect to its antonym: 'blunt' is 'not sharp', and not the other way around. Consequently, if a language has special lexical means to describe the quality 'sharp', it does not necessarily have a simple lexeme for expressing the quality 'blunt' as well. But if there is at least one lexeme covering the field 'blunt' in a language, then there cannot be a lexical gap for its antonym.

6. Discussion: Comparison to previous studies

Previous studies focusing on the properties of lexemes for SHARP and BLUNT are generally limited to one or several closely related languages. This, along with methodologically different approaches explains the discrepancies in the resulting analyses.

Spiridonova (2002) analyses the lexemes *ostryj* and *tupoj* (meaning 'sharp' and 'blunt' respectively) in Russian only. For these adjectives, she distinguishes three classes of direct meanings:

1. usages with functional nouns, characterising their quality according to their efficacy (*ostryj nož* 'sharp knife', *ostraja igla* 'sharp needle', *ostrye kogti* 'sharp claws', *ostrye zuby* 'sharp teeth');
2. usages with nouns denoting objects, marking a special deformation of the end-point (*ostryj špil'* 'pointed spire', *ostryj nos* 'pointed nose', *ostryj kraj* 'sharp edge');
3. usages with nouns that behave like instruments, but the situation also involves an experiencer of the action who reacts to it (*ostryj perec* 'hot pepper').

Spiridonova (2002) also distinguishes three classes of metaphorical meanings, each one being derivative from one of the direct meanings:

1'. sense organs perceived as instruments, hence *ostryj slux* 'keen hearing', *ostryj njux* 'keen sense of smell', *ostroe zrenie* 'keen sight', *ostryj um* 'keen mind';
2'. with abstract nouns (denoting states, processes, and events), *ostryj* means intensity, actualizing the moment or the short span of time in which the state or event is especially strong: *ostryj spor* 'heated dispute', *ostryj moment* 'tense situation', *ostroe vospalenie* 'acute inflammation', *ostroe želanie* 'keen desire';
3'. the third metaphoric meaning involves the figure of the experiencer; the examples are interpreted as stimulating a response: *ostraja kritika* 'sharp criticism', *ostraja satira* 'biting satire', *ostryj yazyk* 'sharp tongue (the quality of being sarcastically witty)'.

The grouping of meanings suggested in (Spiridonova 2002) adequately describes the linguistic behaviour of the Russian adjectives, but it cannot be expanded to the typological data without modifications. As our research shows, in order to capture similarities in the semantics of relevant lexemes in different languages, the first direct meaning in Spiridonova's list should be divided into two subgroups (cutting versus piercing lexemes), while the third meaning should be excluded from the core meanings. There is no typological evidence for the connection between 'pointed shape' and 'heated dispute/tense situation' either (i.e. the connection

between meanings 2 and 2' in Spiridonova's numbering). We have argued based on broader data that this metaphor derives from the functional frames of the SHARP and BLUNT semantic fields.

Another one-language study focusing on the semantics of the field SHARP is found in (Fritz 1995). Working within the action-theoretical semantics framework ("handlungstheoretische Semantik"), Gerd Fritz proposes an analysis of the German word *scharf* accompanied by several short remarks on its intersections with and differences from English *sharp*. He groups the usages of these adjectives around two "centres of gravity". The first is an instrument, such as a knife, with a cutting edge. Such a prototypical instrument triggers an implicational chain for the speaker, according to Fritz: it has a characteristic shape > it functions well > it has a characteristic action > if one is damaged by such an instrument, this will be unpleasant.

The second of the "centres of gravity" refers to human interaction, where the implicational chain is as follows: when an action is intensive (*scharfer Kampf* 'hard or fierce (lit. 'sharp') fight') > it is very effective (*scharfer Argument* 'trenchant argument') > it can be very good (*scharfe Analyse* 'incisive analysis') > it can have an unpleasant effect on the addressee (*scharfe Strafe* 'harsh punishment', *scharfer Urteil* 'harsh judgement').

These two centres of gravity can be associated with the two metaphorical clusters distinguished in our analysis, namely, the functional and the experiential ones. Although with minor discrepancies in details, the division uncovered by Fritz in the German data has typological parallels.

The study of Middle Welsh adjectives by Parina (2016) adopts the same framework and yields comparable results. The adjective *llym* 'sharp' developed both types of metaphors, the first referring to functional perception (for example, collocations with intellectual abilities of a person) and the second one describing experiential perception (for example, the characterization of wind, among others). However, Parina goes beyond the usual means of semantic analysis, such as metaphor and metonymy, and additionally considers culturally specific factors. The complex of the usages of the adjective which are connected with battle, fire, quick wit, and anger is, in her opinion, not culture-independent, but is connected with choleric temperament (cf. Geeraerts & Grondelaers 1995), a notion of the humoral theory that was very influential in the Middle Ages. As for our typological analysis, we do not take such factors into account.

Another typological dichotomy postulated in our analysis is revealed via Swedish data in the study of Vejdemo (2007). A dictionary survey of Swedish near-synonyms *vass* and *skarp* in comparison with English *sharp*, which was augmented with tagging of Swedish and English corpus data, resulted in different

semantic profiles for the three lexemes. While Swedish *vass* and English *sharp* are primarily used for the tactile modality, Swedish *skarp* is associated with the visual modality. This opposition is indirectly reflected in the semantic map of the fields SHARP and BLUNT (Section 3): the rightmost frame (pointed/rounded shape) reflects the visual perception, while the leftmost and the middle ones (cutting/piercing instruments) – the tactile perception.

To our knowledge, the only study to compare data of several unrelated languages is (Goddard & Wierzbicka 2007). This work proposes a definition of English *sharp* designed within the Natural Semantic Metalanguage framework, along with a comparison to the Polish and the Korean correlates. Interestingly, while their definition of English *sharp*[4] reflects the significance of the experiential component, their analysis of the semantics of the Korean lexemes *ppyocokha-* versus *nalkhalop-* results in highlighting the predominance of shape when needle-like objects are opposed to knife-like objects. Note that this distinction is also reflected in the semantic maps proposed in the present study.

In sum, previous studies of lexemes denoting sharpness capture several important regularities. However, only a typological study can result in combining all these regularities together to form a unified semantic system. The present chapter presents an attempt towards such a typological study.

7. Conclusion

In this chapter, we have illustrated the methodology of the Moscow Lexical Typology group through the study of the semantic field SHARP and its antonym BLUNT. Our approach is usage-based, and we pay attention mostly to the contexts in which the target lexemes occur. At the same time, we view contexts as representations

---

4. Goddard & Wierzbicka's analysis is as follows: *This thing is sharp.*

    this thing is like this:
        if a person's hand[m] touches some parts of this thing
        this person can feel something in this hand[m] because of it
        because of this, this person can know something about this thing
        because of this, this person can think like this:
        "if this thing moves in some ways when some parts of it are touching something else, something can happen to this other thing because of it
        it can happen in one moment
        when it happens, some parts of this thing can be inside this other thing because of it
        if this other thing is a part of a person's body,
        this person can feel something bad in this part because of it"

of underlying extra-linguistic situations, which allows us to compare data from different languages.

Our analysis of the fields SHARP and BLUNT demonstrates that such situations (frames, in our terminology) form strictly organized conceptual spaces. The mutual location of frames within these spaces predetermines the possible patterns of lexicalization of semantic fields across languages.

Our frame approach enables efficient analysis of quasi-synonyms within one language, as well as of the so-called translational equivalents – that is, quasi-synonyms from different languages. However, this approach is not restricted to synonyms: as shown, it is also fruitful in describing antonyms. Most importantly, it provides a new type of framework for studying the typology of semantic shifts, focusing on the source meaning and its semantic potential.

Abbreviations

1, 2, 3	the 1st, 2nd, 3rd person		M	masculine
ACC	accusative		N	neutral
ATTR	attributive		NOM	nominative
AUX	auxiliary		PART	partitive
CNV	converb		PFV	perfective
DAT	dative		PL	plural
DET	determiner		POSS	possessive
ERG	ergative		POSTP	postposition
F	feminine		PRS	present
FUT	future		PST	past
GEN	genitive		PTCL	particle
IMP	imperative		REFL	reflexive
INF	infinitive		SG	singular
INS	instrumental		SUP	superlative
LOC	locative			

References

van der Auwera, J. & Plungian, V. 1998. Modality's semantic map. *Linguistic Typology* 2(1): 79–124. https://doi.org/10.1515/lity.1998.2.1.79

Bybee, J. L. & Dahl, Ö. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13: 51–103. https://doi.org/10.1075/sl.13.1.03byb

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

Cysouw, M., Haspelmath, M. & Malchukov, A. L. 2010. Introduction to the special issue «Semantic maps: Methods and applications». *Linguistic Discovery* 8 (1): 1–3. https://doi.org/10.1349/PS1.1537-0852.A.358

François, A. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, M. Vanhove (ed), 163–215. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.106.09fra

Fritz, G. 1995. Metonymische Muster und Metaphernfamilien. Bemerkungen zur Struktur und Geschichte der Verwendungsweisen von "scharf". In *Der Gebrauch der Sprache. Festschrift für Franz Hundsnurscher zum 60. Geburtstag*, H. v. G. Hindelang, E. Rolf & W. Zillig (eds), 77–107. Münster: LIT Verlag.

Geeraerts, D. & Grondelaers, S. 1995. Looking back at anger: Cultural traditions and metaphorical patterns. In *Language and the Cognitive Construal of the World*, J. R. Taylor & R. E. MacLaury (eds), 153–180. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110809305.153

Georgakopoulos, T., Werning, D. A., Hartlieb, J., Kitazumi, T., Peut, L. E. van de, Sundermeyer, A. & Chantrain, G. 2016. The meaning of ancient words for 'earth': An exercise in visualizing colexification on a semantic map. *ETopoi. Journal for Ancient Studies* 6: 418–452.

Georgakopoulos, T. & Polis, S. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2): 1–33. https://doi.org/10.1111/lnc3.12270

Goddard, C. & Wierzbicka, A. 2007. NSM analyses of the semantics of physical qualities: *sweet, hot, hard, heavy, rough, sharp* in crosslinguistic perspective. *Studies in Language* 31(4): 765–800. https://doi.org/10.1075/sl.31.4.03god

Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: Clarendon Press.

Haspelmath, M. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In *The New Psychology of Language, Vol.. 2*, M. Tomasello (ed), 211–242. NJ: Mahwah.

Koptjevskaja-Tamm, M., Divjak, D. & Rakhilina, E. 2010. Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In *New Approaches to Slavic Verbs of Motion*, V. Hasko & R. Perelmutter (eds), 315–341. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.115.18kop

Koptjevskaja-Tamm, M., Rakhilina, E. & Vanhove, M. 2016. The semantics of lexical typology. In *The Routledge Handbook of Semantics*, N. Riemer (ed), 434–455. Oxon/NY: Routledge.

Lazard, G. 1981. La quête des universaux sémantiques en linguistique. *Le Bulletin du groupe de recherces sémio-linguistiques* 19: 26–37.

Levinson, S. & Wilkins, D. (eds). 2006. *Grammars of Space: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486753

Maisak, T. & Rakhilina, E. (eds). 2007. *Glagoly dviženija v vode: Leksičeskaja tipologija* [Verbs of AQUA-Motion: Lexical Typology]. Moscow: Indrik.

Majid, A. & Levinson, S. 2011. The senses in language and culture. *The Senses & Society* 6(1): 5–18. https://doi.org/10.2752/174589311X12893982233551

Majid, A., Gullberg, M., Staden, M. van & Bowerman, M. 2007. How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics* 18(2): 179–194. https://doi.org/10.1515/COG.2007.007

Majid, A., Jordan, F. & Dunn, M. 2015. Semantic systems in closely related languages. *Language Sciences* 49: 1–18. https://doi.org/10.1016/j.langsci.2014.11.002

Parina, E. 2016. The polysemy of llym in Middle Welsh. *Zeitschrift für celtische Philologie* 63: 129–163. https://doi.org/10.1515/zcph-2016-0007

Rakhilina, E. 2010. Verbs of rotation in Russian and Polish. In *New Approaches to Slavic Verbs of Motion*, V. Hasko & R. Perelmutter (eds), 291–314. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.115.17rak

Rakhilina, E. 2010. *Kognitivnyj analiz predmetnyx imen: semantika i sočetaemost'* [Cognitive Analysis of Concrete Nouns: Semantics and Combinability]. Moscow: Azbukovnik.

Rakhilina, E. & Reznikova, T. 2016. A Frame-based methodology for lexical typology. In *Lexico-Typological Approaches to Semantic Shifts and Motivation Patterns in the Lexicon*, M. Koptjevskaja-Tamm & P. Juvonen (eds), 95–130. Berlin, Boston: Mouton De Gruyter. https://doi.org/10.1515/9783110377675-004

Sahlgren, M. 2008. The distributional hypothesis. *Italian Journal of Linguistics* 20: 33–53.

Spiridonova, N. 2002. Prilagatel'nye funkcional'noj semantiki: ostryj i typoj [Adjectives with functional semantics: sharp and blunt]. In *Computational Linguistics and Intellectual Technologies. International Conference « Dialogue 2002 » Proceedings*, 494–499. Moscow: Nauka.

Vejdemo, S. 2007. Sharp, vass och skarp: Semantiska relationer mellan tre perceptionsadjektiv. MA thesis. Stockholms universitet. Institutionen för lingvistik.

A matter of degree?

The domain of wetness in a typological perspective

Tatiana Reznikova, Anna Panina and Victoriya Kruglyakova
HSE University, Moscow / Institute of Oriental Studies, Russian Academy
of Sciences / RANEPA, Moscow

The article studies the domain of wetness in 20 languages. In many of them
the domain features two main words (e.g. German *nass, feucht*; Mongolian *no-
jton, čijgleg*; Moksha *načkə, l'et'kə*) and the difference between them tends to
be described in terms of degree, i.e. 'intensely' versus 'slightly wet'. Typological
analysis shows that in each case the degree of humidity receives a specific in-
terpretation depending on the noun that is being modified, so that the choice
of a particular synonym is based not simply on the quantity of the fluid, but on
the situation as a whole (including the source of moisture, intentional versus
non-intentional event, etc.). We also discuss the additional factors relevant to the
domain in the languages that have more than two words in it, that is, the addi-
tional words with a positive or a negative connotation, or moisture from contact
with a liquid versus moisture absorbed from humid air.

Keywords: physical qualities, semantics of degree, wetness, lexical typology

1. Introduction

This paper discusses the semantic domain WET, which, due to its peculiar structure,
poses a challenge for the methodology adopted by MLexT.

Most of the semantic domains related to physical properties (at least all of the
domains which have been studied by the members of MLexT so far) can be called
"classifying", when the lexemes that constitute the domain collocate only with a
certain subset of nouns, thus dividing all nominal collocates into groups. Languages
with several terms for SHARP, for example, tend to apply them to different kinds
of objects – often there is one term used exclusively for blades with a cutting edge
(knives, scissors, saws, etc.), and another for pointed spikes (needles, spears, arrows,
etc.). For example, French has *tranchant* for blades and *aigu* for spikes: *une lame
tranchante* 'a sharp blade' vs. *une flèche aiguë* 'a sharp arrow', while the reverse

collocations are infelicitous: ˀ*une lame* **aiguë,** ˀ*une flèche* **tranchante** (see Chapter 2 in this volume for details).

Dimensional terms provide another example of a domain where lexemes are distributed across nouns virtually without overlap between the groups of nouns. The subdomain of thickness in some languages features two terms which divide objects conceptualized as having two salient dimensions ("layers") such as ice, cloth or books, vs. those conceptualized as "poles" such as sticks or columns, including also ropes, cables, strings etc.; for example, consider *qaliŋ* for layers vs. *žuan* for poles in Kazakh, cf. **qaliŋ** *mata* 'thick cloth' vs. **žuan** *žip* 'thick rope', which are not interchangeable: **žuan mata,* *qaliŋ žip (see Chapter 5 in this volume).

In both of the above examples each noun can co-occur only with one of the competing lexemes. This tendency is crucial for our approach – our research in various semantic domains shows that meanings of a word can, for the most part, be described and compared cross-linguistically by identifying the possible types of its collocates. Thus, in the case of lexemes denoting physical qualities, it normally suffices, for the purposes of typological comparison, to compare the classes of the collocating nouns.[1] However, for some attributive domains the principle of straightforward distribution of nouns across quality terms, or rather, of quality terms across non-intersecting groups of nouns, does not reveal much about the meaning differences in the quality terms – and wetness is one of such domains.

Most languages have two or three terms for WET, such as *nass* vs. *feucht* in German, *načkə* vs. *l'et'kə* in Moksha, *nureta* vs. *shiketa* vs. *shimetta* in Japanese, or *mokryj* vs. *syroj* vs. *vlažnyj* in Russian. All of these terms can collocate with nouns from the same taxonomic classes – e.g. artifacts or body parts, and, moreover, with the same members of these classes. Instead of a clear-cut distribution characteristic of SHARP or THICK, with different nominal lexemes, we find a mass of overlapping contexts.

This is the case in every language of our sample. English *wet / damp clothes,* German *nasse / feuchte Hände* 'wet / damp hands', French *terre mouillée / humide* 'wet / damp earth', Moksha *načkə / l'et'kə mus'kəpt* 'wet / damp linen', and numerous other pairs suggest that this kind of complicated behavior, not based on the nouns' taxonomic class, is a specific feature of the domain of wetness, which makes both the research into its structure in individual languages and the cross-linguistic comparison thereof much trickier.

It means that for wetness and similar domains we cannot describe distinctions between individual systems by merely equating the typological peculiarity of a

---

1. Note that this principle does not hold for quality terms denoting human properties and states, as in this case the semantic class of the nominal collocates is always the same, i.e. that of human beings.

quality term to a noun class which this term creates. Indeed, it is difficult to find anything resembling a diagnostic context, i.e. a word that would co-occur, e. g. only with 'wet' but not with 'damp' or vice versa.

Still, there is no doubt that the WET terms constitute a semantic opposition of some kind, and their collocations with the same noun will not have identical meanings; the goal of a lexico-typological study is thus to identify such differences in interpretation and to reveal general patterns behind them. The problem with WET-like domains for MLexT is that they appear unyielding to the frame-based method of lexical analysis.

The frame-based method, as described in Chapter 1 of this volume, is based on the analysis of collocations and contexts. In the case of attributive words, we consider the nouns they modify. We are mainly concerned with languages that have more than one lexical term for the domain in question (cf. in the examples above, *aigu* and *tranchant* for SHARP in French, or *qalïŋ* and *žuan* for THICK in Kazakh). Given the fact that each of these attributive terms collocates with a specific group of nouns, the frame-based analysis aims to identify the nouns that appear with each of the attributive terms. In this way, the researcher obtains a list of noun groups that may be lexically opposed to each other within the domain. In the MLexT terminology, such noun groups represent frames (e.g., blades and spikes for the domain of sharpness, layers and poles for the domain of dimensions). The attributive words from different languages are then compared in terms of which frames (i.e., sets of nouns) they can apply to.

Yet, as we show in this study, in the domain of wetness, nouns offer little or no help for differentiating between the quality lexemes, since the collocational behaviours of these lexemes are practically similar. Does this mean that context analysis is of no use for them, and that the frame-based method is not applicable? This paper will show that the contextual analysis, as practiced by MLexT, is capable of dealing with such "abnormal" domains – such domains just require an additional step involving a slightly deeper enquiry into the terms in question and their surroundings.

The study is based on a sample of 20 languages that represent eight families: Indo-European (English, German, Norwegian, Swedish, French, Spanish, Russian, Polish, Lithuanian, Greek, Armenian), Uralic (Hungarian, Khanty, Moksha), Northwest Caucasian (Kabardian), Kartvelian (Georgian), Afro-Asiatic (Hebrew), Sino-Tibetan (Mandarin Chinese), Japonic (Japanese), and Mongolic (Mongolian).[2]

---

2. We would like to express our sincere gratitude to all our informants and experts on these individual languages, who made this research possible: K. Bagdasaryan, K. Böröczki, H. Gérardin, E. Kashkin, L, Kholkina, A. Kozhemyakina, M. Kyuseva, I. Pobedin, E. Rakhilina, D. Ryzhova, E. and S. Sanikidze, O. Shapovalova, E. Shaulsky, A. Skordos, A. Vyrenkova, G. and S. Wollny, S. Yushmanova.

The sample is limited to Eurasia and is biased towards Indo-European languages; however, taking into account the fact that genetically and areally close languages tend to be just as valuable to lexical typology as unrelated ones (see Chapter 1 in this volume), we hope that our research will still reveal the basic typological patterns in the lexicalization of the WET domain. Table 1 summarizes the data on the languages and the lexical units that were considered in this study.

Table 1. Lexical items denoting wetness

Language	Lexical items
English	*wet, moist, humid, damp, dank*
German	*nass, feucht*
Norwegian	*våt, fuktig*
Swedish	*blöt, våt, fuktig*
French	*mouillé, humide*
Spanish	*mojado, húmedo*
Russian	*mokryj, vlažnyj, syroj*
Polish	*mokry, wilgotny*
Lithuanian	*drėgnas, šlapias*
Greek	*vregmenos, mouskemenos, ygros*
Armenian	*t'acʻ, xonav, nam*
Hungarian	*vizes, nedves, nyirkos*
Khanty	*jinki, jinkišǝk, ńar, nivǝŋ*
Moksha	*načkǝ, ľeťkǝ*
Kabardian	*çǝne, vlažne*
Georgian	*sveli, namiani, nesťiani*
Hebrew	*lax, retuv, txuv*
Mandarin Chinese	*shī, cháo, cháoshī*
Japanese	*nureta, shiketa, shimetta*
Mongolian	*nojton, čijgleg*

2.   Amount of moisture and other properties

As we have suggested in Section 1, the domain of wetness allows multiple quality terms to occur with the same nouns. Thus, the difference in interpretation between such collocations must lie not in the properties of objects, but in the very type of wetness. The definitions given in dictionaries and speakers' intuitions alike suggest a plausible explanation: the terms describe different degrees of wetness, i.e. large or small amounts of moisture. This is why even the most respected dictionaries tend to use one of the words in the definitions of the others:

(1) a. English: ***damp*** – *slightly or moderately* ***wet***. (Merriam-Webster)
   b. German: ***feucht*** – *mit Wasser o. Ä. geringfügig durchtränkt oder bedeckt; ein wenig* **nass**
      'slightly infused or covered with water or other liquid; a little **wet**.'
                                                                                    (Duden)
   c. Spanish: ***húmedo*** – *se aplica a las cosas* ***mojadas*** *sólo ligeramente, de modo que, aunque se expriman, no escurre agua de ellas*
      'used about things that are only slightly **wet**, so that even if they are squeezed water does not pour out of them.' (Moliner)

Examples from texts, too, seem to support this interpretation, especially when two terms are used side by side as a contrasting pair:

(2) a. English
      The best way to clean leather is with a ***damp***, *not* ***wet*** *cloth.*
   b. Spanish
      Es mejor que dejes tu cabello algo ***humedo***, *y no* ***mojado***.
      'It is better to have your hair slightly **damp**, and not **wet**'.
   c. Moksha[3]
      Mebel'-t'   *nart-c'ə-saz'*   *af* **načkə**, *a* **l'et'kə**
      furniture-DEF.SG.GEN wipe-FREQ-NPST.3.O.3PL.S NEG wet but damp
      tr'apka-n'a-sə.
      rag-DIM-IN
      'One wipes the furniture with a **damp**, not **wet** rag'.

If the difference between the terms does not go beyond that of degree, then an object that can contain moisture to a greater or lesser extent should theoretically allow of both attributes, i.e. those of higher and lower degree. There is, however, at least one restriction on the usage of the attributes that suggests a more complex distinction: in our sample the lexemes of maximum wetness were rarely or never used to describe air: English ʔ*wet air*, German ʔ***nasse** Luft*, French ʔ*air **mouillé***, Spanish ʔ*aire **mojado***, Russian ʔ***mokryj** vozdux*, Mandarin ʔ***shi** de kongqi*.

   The definition by amount of moisture does not explain why this collocation should not be possible. Humidity in the air is objectively measurable, and it is not beyond human abilities to distinguish, however roughly, its comfortable levels from uncomfortably high ones, so theoretically, we might expect some kind of opposition like the one found in the domain of temperature in such pairs as (moderately) *warm* vs. (too) *hot* (Koptjevskaja-Tamm & Rakhilina 2006; Koptjevskaja-Tamm (ed) 2015, see also Chapter 8 in this volume). Yet, undesirable atmospheric moisture is not expressed by the terms for higher wetness.

---

3. Examples from Moksha were elicited from native speakers.

The case of humid air shows that the distinction between English *wet* and *damp,* German *nass* and *feucht* and similar pairs is not a straightforward matter of degree. Moreover, looking at the contexts in which both terms are possible, we find that the amount of moisture they describe is not the only factor that creates a significant difference in meaning between them.[4] An example of such context is the word for 'hands' combined with 'wet' vs. 'damp' in English and German:

(3) English
    a. *To wash my face with the baking soda, I throw a little bit of baking soda in the palm of my **wet** <u>hand</u>.*   (enTenTen15)
    b. *I listened to my heart thunder in my chest as I clutched the receiver in my **damp** <u>hand</u>.*   (enTenTen15)

(4) German
    a. *Den Teig gut durchkneten und mit **nassen** <u>Händen</u> drei längliche Brote for-men.*   (deTenTen13)
       'Work the dough thoroughly, and form three oblong loaves with your **wet** <u>hands</u>.'
    b. *Eine gute Nachricht gibt es auch für alle, die **feuchte** <u>Hände</u> bekommen, wenn sie eine Rede halten sollen.*   (deTenTen13)
       'There is good news for all who get **damp** <u>hands</u> if they have to make a speech.'

From examples such as these, two contrasting situations emerge as different in more ways than just the amount of moisture the skin. The lexemes that imply a high degree of moisture, i.e. *wet* and *nass*, are used when a person moistens their hands on purpose, e.g. for performing a particular task. The lexemes of low degree, *damp* and *feucht*, describe involuntary, incontrollable perspiration from excitement or fear. The same distribution, 'wet intentionally' vs. 'damp with sweat', is attested in other languages – e.g. Armenian *tʿacʿ* vs. *xonav*, Hebrew *retuv* vs. *lah*, Mongol *nojton* vs. *čijglèg* and others.

    Hands are not the only object for which the terms of wetness vs. dampness describe fundamentally different situations. Consider the following contexts:

(5) English
    a. *Please don't walk on the **wet** <u>floor</u>!*
    b. *Glazed Moroccan tiles are normally water resistant <…>; this means that when cleaning your home using water or in times of flooding during the rainy season the moisture will not leak into the floor material and so you won't have to deal with **damp** <u>floors</u>.*   (enTenTen15)

---

4. Some of the factors underlying the choice between the WET synonyms in English and Russian are discussed in Apresjan et al. (1979) and Apresjan (ed) (2004).

(6)   Russian

    a.   *Sidja*   *na* **mokrom**   *polu*   *(navernoe, kto-to*
        sit:CVB.PRS on wet:LOC.M.SG floor:LOC.SG probably   somebody
        razbil   *akvarium), ona…*   (RNC)
        break:PST.M.SG fish_bowl she.NOM
        'Sitting on the **wet** <u>floor</u> (someone must have broken the fish bowl), she…'

    b.   *Kvartira*   *nosila*   *v sebe*   *zapax*   *nedavnej*
        apartment carry:PST.F.SG in itself:LOC smell:ACC.SG recent:GEN.F.SG
        uborki,   *čistoty,*   **vlažnogo**   *pola*
        mopping:GEN.SG cleanness:GEN.SG moist:GEN.M.SG floor:GEN.SG
        'The apartment had the smell of a recent mopping, of cleanness, of wet (lit. **moist**) <u>floor</u>.'   (RNC)

In the English pair of examples, the *wet floor* has been washed, and the water has not yet dried but is expected to dry up soon; while *dampness* is a persistent, and very undesirable, quality of a floor or an entire room. Russian uses one of its terms for low degree of wetness, *vlažnyj*, to describe a washed floor in (6b), whereas the term for high degree, *mokryj*, mostly refers to the situation of spilled water, as in (6a).

As the examples show, languages may differ in the semantic implications of individual lexemes, but the general principle still holds – a cross-linguistically relevant characteristic for the description of wetness is what caused it. In other words, along with degree of moisture, in the languages examined in this study, the WET terms appear to encode the cause and/or manner of acquiring moisture – that is, whether the moisture appears on the surface from the inside or the outside of an understood "container", whether the situation is temporary or continuous, what process or action brought it about, and so on.

This information, of course, is conveyed not only by the head noun, and this is why the MLexT methodological approach seems at first not to be applicable. Yet, words of wetness can have their own arguments and adjuncts which help differentiate their meaning, e.g. *hands* **damp** <u>with</u> *sweat* or **wet** <u>from</u> *being out in the rain*; however, these extra phrases are never obligatory. Often the information we can look at to further understand the meaning differences is scattered in non-trivial ways throughout the surrounding text, or even requires some reading between the lines – in (6a), for example, 'water' from the broken fish bowl is not explicitly mentioned. This makes it difficult to specify a context in strict, "syntactic", terms.

However, the situational contexts which are united in frames according to the MLexT methodology are semantic, and not syntactic, in the first place. What we call 'frames' are situations, not phrases. It is certainly convenient when the head noun alone is sufficient to identify the type of context for a particular quality term; but in general, it makes no difference how the information is encoded, as long as it can be understood or deduced. So the MLexT method, by and large, still applies.

In view of this, the first step in our research on wetness, similarly to the research on other semantic domains, is to see what core situations, or frames, constitute the domain, which of them can be expressed by the same lexeme in a given language, and which are lexically opposed. Based on this, we look for types of lexical systems in the domain.

3.   Basic systems: Situations and frames

We can take 'humid air' as one of the starting points in our investigation. As discussed above, this situation is invariably expressed by a different word than the one used for surfaces moistened by direct contact with a liquid (cf. 'wet hands after washing'). This allows us to suggest that in languages that make only one lexical opposition in the domain (that is, the binary, 'wet' vs. 'humid' type systems) these situations constitute prototypes for the corresponding subdomains. The other frames are categorized as similar to either 'humid air' or 'liquid-affected surface' and are expressed by the respective terms. All of the binary systems that we have studied have much commonality in the distribution of the terms for 'humid' and 'wet'.

In particular, any objects that have recently been in contact with a liquid are described as 'wet'. No distinction is made for different liquids; **nasse Tischdecke** 'a **wet** tablecloth' in German can mean a tablecloth that has been washed in water, or one over which juice has been spilled. Likewise, it makes no difference whether the source is natural or artificial, and whether the contact is intentional or accidental – *nass* 'wet' can be equally used for a person drenched in the rain, and coming out of a bath or shower.

The term for humidity ('humid'), for its part, metonymically expands its meaning to include not only air with high amounts of moisture, but also the quality of the objects that have absorbed moisture from prolonged exposure to humid air, such as German *feuchte Luft* 'damp air' and *feuchte Kleidung* 'damp clothes'. Another metonymic shift allows the same lexeme to describe time periods and places characterized by humid air – *feuchter Tag* 'a damp day', *feuchter Keller* 'a damp cellar'. A binary opposition similar to the German *nass* 'wet' vs. *feucht* 'damp' can also be found in Spanish (*mojado* vs. *húmedo*), Mandarin (*shī* vs. *cháo/cháoshī*), Polish (*mokry* vs. *wilgotny*), Mongolian (*nojton* vs. *čijgleg*), and Moksha (*načkə* vs. *l'et'kə*).

Different causes that underlie the two contrasting situations correlate naturally with the degree of wetness resulting from them. Indeed, direct contact with a liquid typically produces much more moisture on an object than could be absorbed from the surrounding air. As a result, the two properties, cause and degree, can become conflated – so that 'liquid' is interpreted as causing high degree of moisture, and

'air' is associated with low degree. Hence, degree may become more prominent and take a part in determining which of the two terms should be chosen, particularly in cases where the real cause is neither humid air nor external liquid.

In (4b) we have seen the German term for air humidity / low degree used to describe sweaty hands, even though sweat can be thought of as a liquid. However, unlike situations of a high degree of moisture, in this case the liquid does not come from outside, but is secreted inside the body.

Another object in which a low degree of moisture is associated with internal moisture is soil. Places which are habitually damp are described by the term for low degree, while the term for high degree is usually reserved to denoting the effects of moisture coming from the outside, e.g. when the ground is wet with rain:

(7) Polish
   a. *Proszę się   wybrać na podmokłą   łąkę   czy*
       please REFL choose on swampy:ACC.F.SG meadow:ACC.SG or
       brzeg   potoku.   Tam, gdzie jest   żyzna
       bank:ACC.SG creek:GEN.SG there where be:PRS.3SG fertile:NOM.F.SG
       i   wilgotna   ziemia.   Znajdzie   tam   pani
       and moist:NOM.F.SG soil:NOM.SG find:FUT.3SG there lady
       walerianę.   (NKJP)
       valerian:ACC.SG
       'You can go to a swampy meadow or to a creek bank – there where the soil is fertile and moist. You will find valerian there'.
   b. *Ściemniło   się   i   zaczął   padać deszcz,*
       get_dark:PST.N.SG REFL and begin:PST.M.3SG fall:INF rain:NOM.SG
       krowy   poszły   do obór.   Przykucnęłam   pod
       cow:NOM.PL   go:PST.3PL   to cowshed:GEN.PL crouch:PST.F.1SG under
       krzakiem, mokra   ziemia   nie pozwalała   mi
       bush:INS.SG wet:NOM.F.SG ground:NOM.SG not allow:PST.F.3SG I:DAT
       się   położyć.   (NKJP)
       REFL put:INF
       'It grew dark and began to rain, the cows went to the shed. I crouched under a bush, I could not lie down due to wet ground'.

Finally, if there is a conflict between the degree and the cause, the degree can override the cause. In the examples like the partially dry hair in (2b) and washing in (8), the term for air humidity / low degree is used to describe objects that have been immersed in water, but enough time has elapsed for most of moisture to have dried off their surface.

(8) Moksha

 šobdava-stə mus'kəm-s' kos'k-i, no son n'ingə l'et'kə.
 morning-EL bed_linen-DEF.SG dry-NPST.3SG but it   still   moist
 'The bed linen has been drying since morning, but it's still moist'.

The distribution of some situations in a binary, 'wet' vs. 'humid' type system is summarized in Table 2.

Table 2. Distribution of situations in a binary system

Wet / high degree lexeme	Humid / low degree lexeme
solid object, after immersion in water	humid air
cloth, from spilled water	cloth stored in a damp place
cloth, directly after washing	cloth that has not completely dried yet
time period with lots of rain	time period with high humidity
	place with humid air
hands after dipping them in water	sweaty hands
ground after the rain	habitually damp ground

It must be noted that cause of moisture and amount of moisture are not the only properties of a situation that may correlate with each other. Similarly to other physical qualities discussed in this volume, such as sharpness or smoothness, the choice among terms for wetness can to some extent reflect the mode of perception, namely, visual vs. tactile. Direct contact with a liquid normally affects the appearance of an object much more than when it is affected by absorbing the ambient moisture. In the former case, some of the liquid stays on the surface where it can be seen, and if it seeps inside, the surface may temporarily change its color. When moisture is accumulated from the air, the amount is generally too insignificant to produce visible changes, and so its presence can usually be discovered only by touch (for Russian, cf. Apresjan (ed) (2004: 539–540)). Thus, in German (see (9)), we find the term for high degree of wetness in the context of either tactile or visual perception, while the term for low degree implies tactile perception only.

(9) German
 a. *Insbesondere sind es die **nassen** Blätter, auf denen dann grelle Lichtflecken zu sehen sind.* (deTenTen13)
  'In particular, it is the wet leaves, on which bright light spots can then be seen'.
 b. ***Feuchte** Blätter klebten an meiner Haut und ich atmete die kalte, modrige Erde ein.* (deTenTen13)
  'Moist leaves were sticking to my skin and I inhaled the smell of cold, moldy earth'.

Apart from the mode of perception, another property of the situation which correlates with terms for amount of moisture is how long the situation persists. An object that has become wet from contact with a liquid (has been washed, left in the rain, had something spilled on it, etc.) normally becomes dry again after some time, meaning that this characteristic is temporary. Ambient or internal moisture is typically an inherent property of an object (a marshy spot, a damp cellar, etc.), and does not wear away by itself. As a result, terms for a high level of wetness are associated with transient states, while dampness tends to be conceptualized as habitual.

An interesting consequence of this fact can be seen in the syntactic class of the respective terms. Most other domains of physical qualities deal with more permanent attributes, such as 'red' or 'narrow', and the terms which comprise them are typically adjectives in languages that have adjectives as a word class. The domain of wetness is different in this respect; the state that it expresses can be temporary, and it seems to be not coincidental that in some languages, alongside adjectives, we encounter verbs or verbal derivatives denoting wetness.

In some lexical systems the wetness domain is populated exclusively with verbs (such as Japanese, with its verbal derivatives *nureta* from *nureru* 'to get wet', *shiketa* from *shikeru* 'to get damp', and *shimetta* from *shimeru* 'to get damp'); others have only adjectives (e.g. German *nass, feucht*; Moksha *načkə, l'et'kə*). Mixed systems also exist, where some of the wetness terms are purely adjectival, while others are either verbs or forms derived from verbs. In our sample, these mixed systems reveal a remarkable pattern: the term for high degree of moisture – which is a transient state – will be verbal, while the term for low degree of moisture – which is a more persistent attribute – will be an adjective. In Spanish, for example, the high-degree term *mojado* 'wet' is a participle of the transitive verb *mojar* 'to moisten', and the low-degree term *húmedo* 'damp' is an adjective.

The contrast in duration is also reflected in the terms' grammatical behavior in Mandarin Chinese. The domain is divided between two terms – the high-degree *shī* and the low-degree *cháoshī*, both of which can appear in the attributive position, e.g. **shī** *yīfu* '**wet** clothes', **cháoshī** *de yīfu* '**damp** clothes'. The predicative construction, however, demands a transient condition rather than a permanent state, and it is remarkable that only the high-degree term is possible in this construction: *yīfu* **shī** *le* 'the <u>clothes</u> became **wet**', **yīfu cháoshī le* 'the <u>clothes</u> became **damp**'.

Summarizing the discussion above, the most simple, binary systems consist of two lexemes, each encompassing a group of situations. One group is formed around the situations where an object has recently been in contact with external liquid, while the central frame for the other group corresponds to humid air and the states caused by exposure to it. This basic opposition between the two terms is supplemented by a number of other contrasting properties of the situations, most

prominently the high or low amount of moisture, as well as the duration of the state and the primary mode of its perception. The next section will explore the structure of lexical systems where the wetness domain is covered by more than two terms.

4. Richer systems

As we have seen in Section 3, when the basic systems divide the domain of wetness in half around their respective core concepts, they conflate several types of situations (cf. 'cloth stored in a damp place' and 'cloth that has not completely dried yet'). Systems that have more than two terms lexicalize additional semantic properties and distinctions. This section looks at the semantic properties that can further structure the domain of wetness.

For the situation of high degree of wetness, which results from contact with a liquid, the nature of the liquid can be important. One example is found in Hungarian, where water is opposed to other liquids: the language has a specific term, *vizes* 'wet with water', derived from *viz* 'water'. Similar amounts of any other liquid are described with the adjective *nedves* 'wet'. Thus there are two ways to translate an English sentence like *Don't put anything on the wet table* into Hungarian – either *vizes* or *nedves* must be chosen depending on the liquid, as seen in (10).

(10) Hungarian
    a. *Ne tegyél   semmit a   **vizes**   asztalra.*
        not do:SBJV.PRS.2SG nothing DEF wet_with_water table:SUBL.SG
        'Don't put anything on the wet table {wet with water}'.
    b. *Ne tegyél   semmit a   **nedves** asztalra.*
        not do:SBJV.PRS.2SG nothing DEF wet   table:SUBL.SG
        'Don't put anything on the wet table {wet with something other than water}'.

Terms for low amount of moisture also can lexicalize additional semantic distinctions. As we suggest above, the core situation for this "low" part of the domain is humid air and its effect. Binary systems use the same term to refer to moisture caused by contact with a liquid, provided the amount is sufficiently low – as when the object has had enough time to partially dry. Rich systems do not need to place these two situations into the same class, but can make the source of moisture a basis for lexical distinction. This distinction is found in Georgian – *nest'iani* is the term for dampness caused by atmospheric moisture, and *namiani* for the state of not being quite dry after contact with a liquid, as illustrated in (11).

(11) Georgian[5]

    a.   *is*   *t^{h}et^{h}reuls*   *sardap^{h}ši inaxavda,*   *amiťom sť'umrebs*

         she   bed_linen:DAT cellar:LOC keep:IMPF.3SG so   guest:PL.DAT

         *q'oveltvis **nesť'ian** t^{h}et^{h}reulši*   *uc'evdat*   *dadzineba.*

         always   damp   bed_linen.LOC have_to:IMPF.3PL sleeping:NOM

         'She always stored her bed linen in the cellar, so that her guests had to sleep on damp bed sheets'.

    b.   *recxvis*   *šemdeg k'argad gaašret^h*   *t^{h}et^{h}reuli,*   *ar*

         washing:GEN after   well   dry:IMP.PL bed_linen:NOM not

         šeidzleba   ***namiani** t^{h}et^{h}reulis*   *k'aradaši*   *šenaxva*

         it_is_possible moist   bed_linen:GEN wardrobe:LOC saving:NOM

         'Dry the bed linen properly after washing, you shouldn't put the moist bed linen into the wardrobe'.

One more source type for which there often exists a special term is sweat. Many languages describe its presence with a derivate such as the English *sweaty*, cf. Latvian *sasvīdis* from the verb *svīst* 'to sweat', Dutch *zweterig* from the noun *zweet* 'sweat', Russian *potnyj* from *pot* 'sweat', and others. It is important to note that in some languages the specific term for being 'wet with sweat' can be preferred for the relevant situation, and this term is not interchangeable with the general term for wetness. This is the case in Georgian, where none of *sveli, namiani*, or *nesť'iani* can be used instead of *op^{h}liani* 'sweaty' (from *op^{h}li* 'sweat'), as illustrated in (12).

(12) Georgian

     me vt^{h}xovdi   *dacʼqʼnarebuliqʼo,*   *xelši*   *meč'ira*   *misi*

     I   request:IMPF.1SG calm_down:PQP.3SG hand:LOC hold:AOR.1SG her

     ***op^{h}liani** xeli*

     sweaty   hand:NOM

     'I asked her to calm down, held her damp hand in mine'.

Besides the physical properties of the situation, some systems have terms that express whether the situation is subjectively evaluated as undesirable. Presence of moisture can mean that an object's function is damaged, as with lumpy salt, firewood that does not burn properly, linen too damp to sleep on, etc. Some languages have lexical means to emphasize this negative effect. Pairs of terms for 'damp' with different evaluation can be found in Russian – *syroj* (negative) and *vlažnyj* (neutral), cf. Apresjan (ed.) (2004:.541), Hebrew – *taxuv* (negative) and *lax* (neutral), or Japanese – *shiketa* (negative) and *shimetta* (neutral), as illustrated in (13).

---

5. Examples from Georgian were elicited from native speakers.

(13) Japanese
    a. **Shiketa** <u>shio</u> no, kantan saisei   hōhō   (jaTenTen11)
        damp   salt GEN simple restoration method
        'An easy way to dry out (lit. restore) **damp** <u>salt</u>.'
    b. *Kawaita shio to* **shimetta** <u>shio</u> de wa, jakkan   omosa ga
        dry   salt and damp   salt INS TOP somewhat weight NOM
        kotonarimasu   (jaTenTen11)
        differ
        'The weight of dry and **damp** <u>salt</u> is slightly different.'

Interestingly, evaluation often appears to correlate with temperature. Negative evaluation is usually associated with cold, as in the English adjective *dank* 'unpleasantly damp, often chilly', cf. *dank and cold* vs. ʔ*dank and hot,* see also (Apresjan et al. 1979:.481–482). The neutral term is typically used to describe the situations when the air temperature is high or normal, as in the Russian examples in (14).

(14) Russian
    a. *Xolodnyj*   **syroj**   <u>vozdux</u>   *to i delo*
        cold:NOM.M.SG damp:NOM.M.SG air:NOM.SG at_times
        obdaval   *iz*   *černogo*   *provala.*   (RNC)
        pour_over:PST.M.SG from black:GEN.M.SG ravine:GEN.SG
        'At times gusts of cold **damp** <u>air</u> blew from the black ravine.'
    b. *Zapadnye*   *vetry*   *prinosjat*   *zimoj*   *na*
        westerly:NOM.PL wind:NOM.PL bring:PRS.3PL winter:INS.SG on
        bol'šuju   *čast'*   *Evropy*   *teplyj,*
        bigger:ACC.F.SG part:ACC.SG Europe:GEN warm:ACC.M.SG
        vlažnyj   *(# syroj)*   <u>vozdux</u>   *Atlantiki.*   (RNC)
        humid:ACC.M.SG damp:ACC.M.SG air:ACC.SG Atlantic:GEN
        'In winter, westerly winds bring warm, **humid** Atlantic <u>air</u> to the most part of Europe.'

The presented overview has covered the basic oppositions which we have encountered in the domain of wetness in our typological sample. The set of the core situations outlined by these oppositions can be used to create a semantic map – a tool which enables visualizing the scope of meaning of individual terms and comparing them across languages.

5. Constructing a semantic map

Using the lexical oppositions observed in our data we can construct a semantic map of the basic layout of the domain (Figure 1). The similarities and differences among lexical systems in individual languages can then be plotted on this map.

'wet with water'	'damp because of humid air'	'humid (about cold air)'
		'humid (about hot air)'
'wet with other liquid'	'not yet dry after contact with liquid'	
	'damp with sweat'	
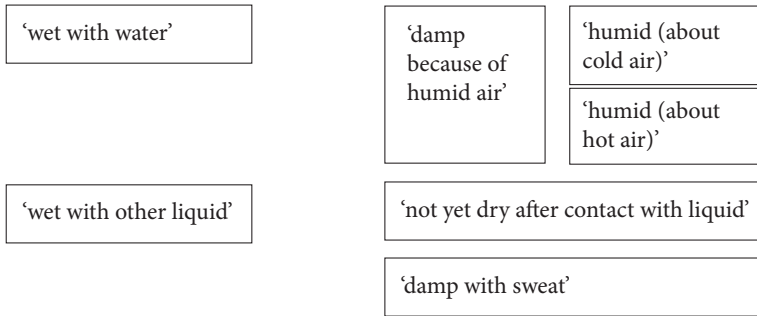
Figure 1. Basic semantic map of the WET domain

The map is clearly divided in half: the situations of intense wetness are grouped on the left, and the situations of low-level humidity are on the right. Every frame is involved in some of the situations of lexical choice which were discussed in Sections 3 and 4.

The only parameter that needs to be mentioned separately is evaluation. Technically, any situation in the domain could be thought of as undesirable, so this parameter is not reflected in Figure 1. Our typological sample, however, strongly tends to ascribe negative evaluation to the dampness caused by humid air (e.g. linen that has been stored in a damp room). This tendency is not without extra-linguistic basis, as humid air is hardly ever applied deliberately; normally when something needs to be moistened, a liquid is used. Furthermore, humidity is difficult to protect from and not visibly noticeable, except for its effect on other objects, which often renders things unusable (e.g. damp matches or mouldy bread). It is rather natural, then, that such situations should often warrant a negatively loaded term.

The correlation between evaluation and cause is a good example for demonstrating that many of the oppositions discussed in Section 4 are not independent, cf. (Rakhilina & Reznikova 2016) and Chapter 1 in this volume; for an alternative approach, see (Lehrer 1974). Languages do not develop words for every possible combination of the factors identified in this paper, or for random combinations; rather, the factors combine to describe meaningful situations, and these situations are what we call frames. Accordingly, our basic map does not use any separate notation for evaluation, and negative terms for 'dampness' in specific languages are represented by the frame 'damp because of humid air'.

Lexical systems in which such a negative term exists show some variation in whether it can be applied to air or not. In some languages, such as Russian, air can be described as unpleasantly damp (*syroj vozdux*). As discussed in Section 4, this usage commonly implies low ambient temperature. In others, such as Armenian and Hebrew, the negative term is limited to objects accumulating undesirable moisture from the air, while air itself allows of only the neutral term for low-level humidity. We represent this difference in scope by positing separate frames for cold and warm air. Languages in which the negative term can be applied to air are the ones in which the opposition is lexicalized.

The next section will show how individual lexical systems distribute their vocabulary over the framework presented in the basic map.

6.   Lexicalization of the domain of wetness in individual languages

The simplest pattern of lexicalization is exhibited by the most basic, binary systems, which classify all frames by the associated amount of moisture. The semantic map in Figure 2 shows one term in the left-hand half of the map, denoting the situations of (intense) wetness, and the other term in the right-hand half which corresponds to (low-level) humidity. This distribution, illustrated in Figure 2, is attested in Spanish (*mojado* vs. *húmedo*), German (*nass* vs. *feucht*), Polish (*mokry* vs. *wilgotny*), Moksha (*načkə* vs. *ĺet'kə*), Mandarin (*shī* vs. *cháo/cháoshī*), and Mongol (*nojton* vs. *čijgleg*).
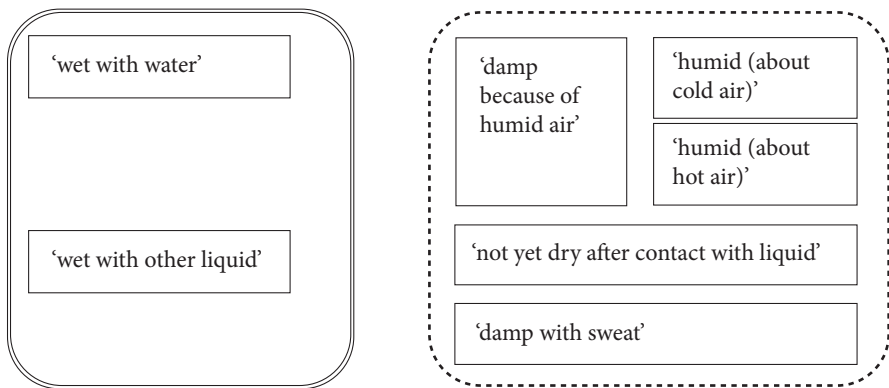
Figure 2. Semantic map of the WET domain: Binary systems

A slightly more complicated system emerges in the languages that introduce the additional parameter of evaluation; the result, as we remember, is three separate terms: wetness vs. neutral humidity vs. negative dampness.

Russian can serve as an example of such a triple opposition, shown in Figure 3. Similarly to the binary systems, the high degree of wetness is covered by a single term, the adjective *mokryj*, which can describe the effect of recent contact with a liquid. Low-degree humidity is expressed by two adjectives – *syroj* and *vlažnyj*. Of them, *syroj* carries a negative connotation and describes cold humid air and objects that have accumulated undesirable moisture from exposure to it, while *vlažnyj* covers the remaining frames.

It must be noted that dampness with sweat, despite its negative connotations, does not fall into the scope of *syroj* and its analogs in other languages. A possible explanation may be that the negative term for dampness typically implies low temperature, while sweat is associated with heat or fever.
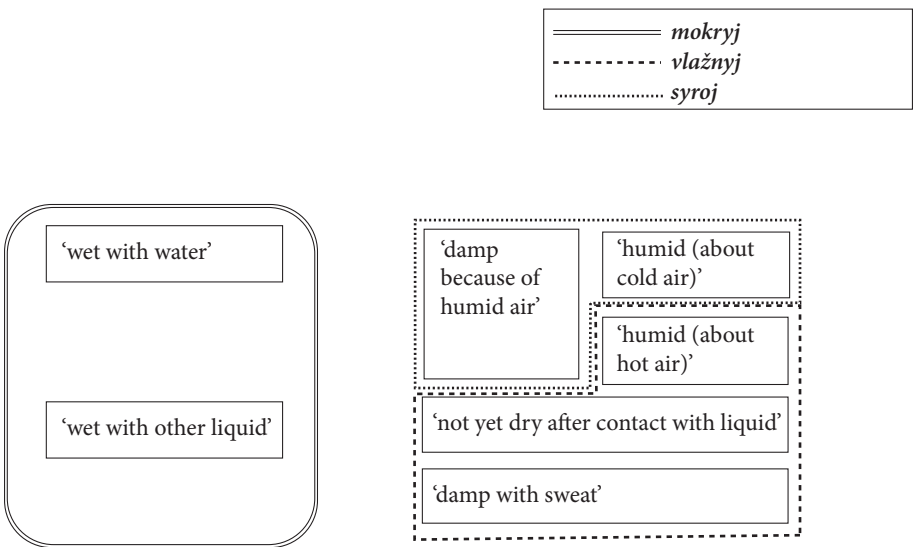
Figure 3. Semantic map of the WET domain: Russian

A slightly different variety of ternary systems is found in the languages that do not use the negatively marked term for air, as mentioned in Section 5. Armenian, like Russian, has three terms: *t'ac'*, similarly to Russian *mokryj*, covers recent contact with a liquid; but the terms for dampness are distributed differently. The negative term, *nam*, describes objects which have accumulated moisture from the air (this also includes rooms with damp air). The neutral term, *xonav*, is used for air irrespectively of temperature and for all the remaining frames with low amounts of moisture.

The system is shown in Figure 4.
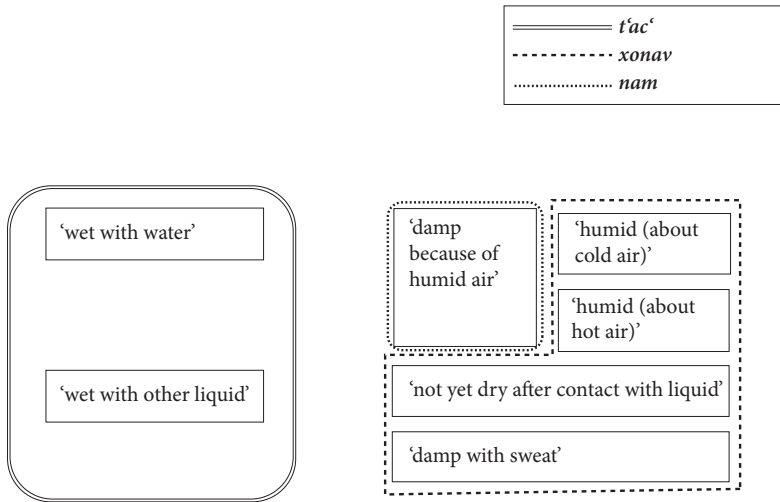
Figure 4. Semantic map of the WET domain: Armenian

Other languages make even more distinctions within the low-intensity half of the domain. Georgian has four terms, but the whole left half of the map is still covered by a single adjective of intense wetness, *sveli*. On the right we find three terms for moisture from varying sources: air vs. sweat vs. other liquids. The adjective *nest'iani* can be applied to objects which have absorbed moisture from the air, as well as to air itself, either warm or cold. Slight residual dampness after contact with a liquid, when the object has had time to partially dry, is described by *namiani*. Finally, *opʰliani*, derived from the term for sweat, describes sweaty objects – mostly body parts; none of the other adjectives, *sveli, nest'iani* or *namiani*, can be used in the context of, for example, sweaty palms.

Figure 5 illustrates this system.

An additional opposition within the high-degree wetness subdomain was found only in one language in our sample, Hungarian, which has a dedicated term for wetness with water – *vizes*.

One more unique characteristic of the Hungarian system is that not all of the adjectives in its wetness domain make the distinction between high and low amount of moisture. Objects wet with something other than water are described by *nedves*, which is the term that also covers various situations of dampness (including water, if it has almost dried). It is possible that Hungarian treats all other liquids as somehow "less wet" than water, making *vizes* the only real term for wetness. Alternatively, degree may not in fact affect the set-up of the wetness domain in Hungarian. Its main term is *nedves*, which is unspecified as to the degree, while *vizes* exists on
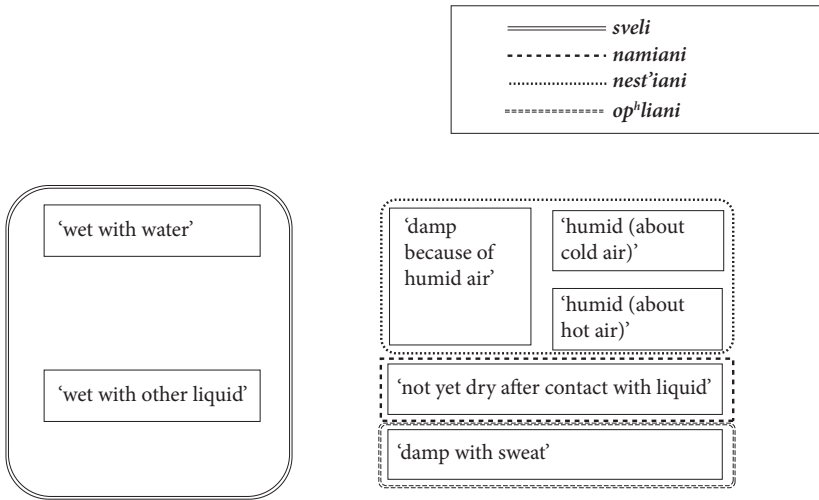
Figure 5. Semantic map of the WET domain: Georgian

the periphery of the domain. The third term is the negative lexeme *nyirkos*, which describes unwanted moisture absorbed from air.

Hungarian has one more term to describe damp air *párás*, whose central meaning is not just 'moisture in the air' but 'visible mist'. The word is used in collocations with transparent objects such as lenses or windows to indicate that they are fogged. Therefore we consider *párás* to belong to the adjacent domain, 'misty', and do not include it in the map of the Hungarian domain of wetness, shown in Figure 6.
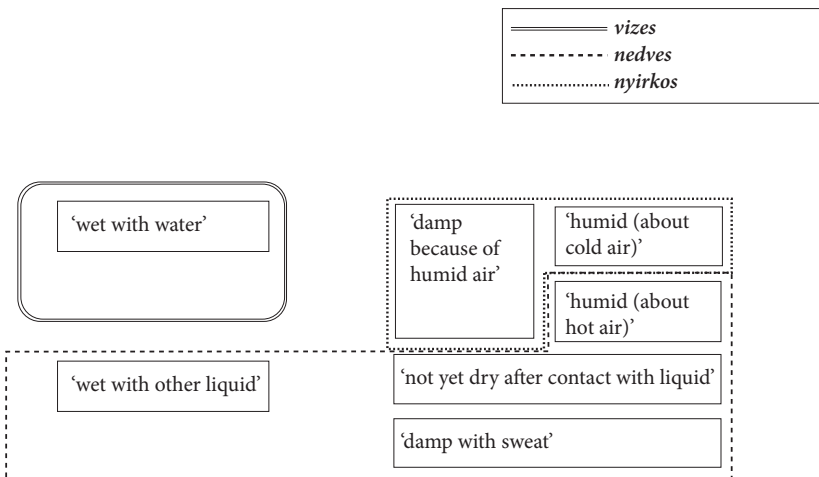
Figure 6. Semantic map of the WET domain: Hungarian

7.   Conclusion

In this paper we have presented a typological study of the semantic domain of wetness. The lexicographic tradition of several languages (cf. (1)) suggests that the terms of the domain form an opposition of degree, i.e. the lexical choice depends on the amount of moisture: intense 'wetness' vs. slight 'dampness'.

If degree was the only difference between the terms corresponding to 'wet' and 'damp', they should be interchangeable in any context (i.e. with any noun) without a significant difference in meaning. Instead, our data from different languages shows that often it is impossible to replace a term by its near synonym. Either the synonym cannot be used in a given context at all, or it completely changes the understanding of the circumstances under which moisture was acquired. Essentially, each term is associated with a number of situations of wetness or dampness, and the distribution of these situations among the terms varies from language to language, even though some overall tendencies appear to be universal.

All languages in our sample use the term associated with 'high degree of wetness' for objects soaked in water and the term related to 'low degree of wetness' for humid air, but other situations show certain variation. Larger amounts of liquids other than water may be described by the high-degree term, as in German, Mandarin Chinese and others, or by the low-degree term, as in Hungarian. See also such pairs as (5a-6b), where a recently washed floor is described as *wet* in the English example, but a similar context in Russian permits a low-degree term. These subtle differences between languages make it clear that degree of wetness is not sufficient to explain the full range of differences in meaning, and frames are necessary for a comprehensive analysis of the vocabulary of wetness.

In conclusion let us point out that the way in which 'amount' of a feature tends to correlate with other characteristics is not limited to the domain of wetness. Degree is seldom realized as the simple opposition of 'less vs. more', and the tendency for words whose meanings differ only in respect of degree to acquire additional semantic nuances is probably quite widespread in human language, manifesting itself on different structural levels. Historical studies show that the grammatical category of degree is unstable and often evolves into other grammatical meanings, such as aspect, cf. (Heine & Kuteva 2002; Plungian 2011). Similarly, in word-formation it is common for diminutives, in addition to their prototypical meaning of small size, to develop such meanings as 'young animal', 'unit of a substance', or 'imitation', see (Jurafsky 1996; Spiridonova 1999). Our data can be interpreted as a manifestation of the same phenomenon on the lexical level.

Abbreviations

1, 2, 3	the 1st, 2nd, 3rd person		LOC	locative
ACC	accusative		M	masculine
AOR	aorist		N	neutral
CVB	converb		NEG	negation
DAT	dative		NOM	nominative
DEF	definite		NPST	non-past
DIM	diminutive		O	object conjugation
EL	elative		PL	plural
F	feminine		PQP	pluperfect
FREQ	frequentative		PRS	present
FUT	future		PST	past
GEN	genitive		REFL	reflexive
IMP	imperative		S	subject conjugation
IMPF	imperfective		SBJV	subjunctive
IN	inessive		SG	singular
INF	infinitive		SUBL	sublative
INS	instrumental		TOP	topic

References

Apresjan, Ju.D. (ed). 2004. *Novyj ob"jasniteľnyj slovar' sinonimov russkogo jazyka* [New Explanatory Dictionary of Russian Synonyms]. Moscow: Jazyki slavjanskoj kuľtury.

Apresjan, Ju.D., Botiakova, V. V., Latysheva, T. E., Mosiagina, M. A., Polik, I. V., Rakitina, V. I., Rozenman, A. I. & Sretenskaia, E. E. 1979. *Anglo–russkij sinonimičeskij slovar'* [English–Russian Dictionary of Synonyms]. Moscow: Russkij jazyk.

Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613463

Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language 72*(3): 533–578. https://doi.org/10.2307/416278

Koptjevskaja-Tamm, M. (ed). 2015. *The Linguistics of Temperature* [Typological Studies in Language, 107]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.107

Koptjevskaja-Tamm, M. & Rakhilina, E. 2006. « Some like it hot »: On the semantics of temperature adjectives in Russian and Swedish. *STUF (Sprachtypologie und Universalienforschung)* 59(2): 253–269. https://doi.org/10.1524/stuf.2006.59.3.253

Lehrer, A. 1974. *Semantic Fields and Lexical Structure*. Amsterdam: North-Holland, New York: American Elsevier.

Plungian, V. A. 2011. *Vvedenie v grammatičeskuju semantiku. Grammatičeskie značenija i grammatičeskie sistemy jazykov mira* [Introducing Grammatical Semantics: Grammatical Values and Grammatical Systems in the World's Languages]. Moscow: RGGU.

Rakhilina, E. V. & Reznikova, T. I. 2016. A frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–129. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110377675-004

Spiridonova, N. F. 1999. Russkie diminutivy: problemy obrazovanija i značenija [Russian diminutives: formation and meaning]. *Izvestija AN, SLJa* 58(2): 13–22.

Dictionaries and corpora

deTenTen13 – German Web Corpus on the Sketch Engine platform <https://www.sketchengine.eu/detenten-german-corpus/> (20 August 2018).

Duden – Duden Online-Wörterbuch <http://www.duden.de> (20 August 2018).

enTenTen15 – English Web Corpus on the Sketch Engine platform <https://www.sketchengine.eu/ententen-english-corpus/> (20 August 2018).

jaTenTen11 – Japanese Web Corpus on the Sketch Engine platform <https://www.sketchengine.eu/jatenten-japanese-corpus/> (20 August 2018).

Merriam-Webster – Merriam-Webster Dictionary <http://www.merriam-webster.com> (20 August 2018).

Moliner – Moliner, M. 1998. *Diccionario de uso del español*. 2nd ed. Barcelona: Editorial Gredos.

NKJP – Narodowy Korpus Języka Polskiego <http://nkjp.pl> (20 August 2018).

RNC – Russian National Corpus <http://ruscorpora.ru> (20 August 2018).

Quality as a two-place predicate
The typology of FULL and EMPTY

Anna Panina and Maria Tagabileva

Institute of Oriental Studies, Russian Academy of Sciences /
Independent researcher

This chapter presents a typological description of the semantic domains EMPTY
and FULL. While the majority of other predicates of physical qualities have a
single argument (the qualified object), the situations of fullness and emptiness
imply two arguments – the container and the contents. Such argument structure
influences the syntax of the predicates of fullness and emptiness and, to some
extent, the oppositions within these domains. In this paper, EMPTY and FULL are
examined in relation to CLEAN, BARE, SOLID and several other adjacent seman-
tic domains, with particular attention being given to metaphors and semantic
extensions.

Keywords: qualitative adjectives, full, empty, container, lexical typology

1. Introduction

In this chapter we take a look at the domains EMPTY and FULL, where the presence
or absence of contents is stated as a quality of the container.

We employ the frame-based approach introduced in detail in Chapter 1. Prior
work on these domains is limited, yet a thorough study of the adjectives meaning
'empty' in Slavic languages, modern and extinct, can be found in (Tolstaya 2008);
besides, some of our earlier findings were published in (Tagabileva & Kholkina
2010) and (Tagabileva et al. 2013).

One prominent feature of the predicates of emptiness and fullness is that they
feature two semantic arguments, and overt realization of both arguments is often
possible (*X is full/empty of Y*). Predicates of physical qualities, especially adjectives,
tend to have only one argument, the noun which they modify; however, it has been
shown that adjectives of qualities may range on the scale from one-place predicates
to predicates with multiple arguments (Kustova 2006; Bhat 1999). For example,

among the domains discussed in this volume, WET has two semantic arguments: a wet object implies the presence of some liquid, as in *wet <u>with tears</u>*. The domains EMPTY and FULL also provide a few more examples of two-place predicates which will be discussed in the subsequent sections of this chapter.

Our sample is rather small: it includes only eight languages for which some preliminary research of the domains EMPTY and FULL has already been conducted by our colleagues or students.[1] Despite its arbitrary nature, our sample represents several families (Indo-European, Finno-Ugric, Sino-Tibetan, and Altaic) while at the same time two of the languages (Russian and Serbian) form a closely related pair. The languages of the sample are English and Spanish (regardless of their regional varieties), Russian, Serbian, Khanty, Mandarin Chinese, Japanese and Korean.

For both of the domains under examination (EMPTY and FULL), we are going to discuss, firstly, the physical situations central to these domains (Sections 2.1 and 3.1, respectively), then the adjacent domains and the vocabulary which is either shared by all of them or which is competing across them (Sections 2.2 and 3.2), and ultimately, the most prominent semantic extensions (Sections 2.3 and 3.3). The chapter concludes with preliminary semantic maps summing up the structure of the domains as attested in our sample (Section 4).

2.   Empty

2.1   The main (physical) sense

The notion of physical emptiness encompasses two core situations which are often lexicalized separately: 'hollow object' vs. 'container without contents.' This opposition closely resembles the oppositions found in some other domains presented in this volume – the ones that can be roughly generalized as 'form vs. function'. Here, form is represented by 'hollow', and function by 'empty of contents': container X, natural or artificial, is 'empty' of contents Y when Y is expected to be located within X, but is missing. A non-container object X is 'hollow' when some of X's mass, located under the surface of X, is missing.

The following languages of our sample have separate terms for 'empty containers' vs. 'hollow shape': English (*empty* vs. *hollow*), Russian (*pustoj* vs. *polyj, pustotelyj*), Serbian (*prazan* vs. *šupalj*), Spanish (*vacío* vs. *hueco*), and Mandarin Chinese (*kōng* vs. *kōngxīn*):

---

1.   The authors thank L. Kholkina for the data on Mandarin Chinese, N. Muraviev on Khanty, E. Rudnitskaya and A. Sorokina on Korean, and A. Spesivtseva on Spanish.

(1) Spanish

    a. *¿Es   cierto que dej-ar   el   plato **vací-o**   es   de*
       be.3sg true   REL leave-INF DEF.M.SG plate empty-M.SG be.3sg of
       mal-a   educación?   (esTenTen18)
       bad-F.SG education
       'Is it really bad manners to leave your plate **empty**?'

    b. *En el   tronco **huec-o**   de un   manzano silvestre…*
       in   DEF.M.SG trunk hollow-M.SG of   INDF.M.SG apple_tree wild
       'In the **hollow** trunk of a wild apple tree…'   (CE)

(2) Serbian

    a. *Zašto se   gospodi čini   da   je*
       why   REFL Lord   seem:PRS.3SG CONJ be:PRS.3SG
       *bolja   puna   nego **prazna**   čaša?*
       good:COMP.NOM.F.SG full:NOM.F.SG than empty:NOM.F.SG glass:NOM.SG
       'So why, o Lord, do we feel that a full cup is better than an **empty** one?'
                                                              (BCS)

    b. *Ovaj   zid   je   **šupalj**.*   (Glosbe_SR)
       this:NOM.M.SG wall:NOM.SG be:PRS.3SG hollow:NOM.M.SG
       'This wall is **hollow**.'

(3) Mandarin

    a. *Qǐng gěi wǒ yī gè **kōng** bēizi.*
       please give I   one CLF empty glass
       'Give me an **empty** glass.'

    b. *Lǎo huáishù **kōngxīn** le.*   (CCL)
       old sophora hollow   MOD
       'The old sophora tree has become **hollow**.'

In all of the surveyed languages which have a term for 'hollow', its use is limited. It can be replaceable by the term for 'empty of contents', and/or derived from it, but not vise versa. The Mandarin *kōngxīn* 'hollow', lit. 'empty-centered', is a compound derived from *kōng* 'empty of contents'. Japanese has a Chinese loan term for 'hollow', *kūdō*, but there are examples where *kara* or *karappo* 'empty' is used, such as (4b); an older term for 'hollow', *utsuro*, is mostly used in metaphors. The Russian adjective *pustotelyj* lit. 'empty-bodied', a near synonym of *polyj* 'hollow', is derived from *pustoj* 'empty', and there is a tendency for *pustoj* to take over situations covered by *polyj* 'hollow', so that the latter has been driven almost completely out of informal speech, as in Example (5).

(4) Japanese
  a. *Miki no naka ga **kūdō** / **kara** ni natte-iru.*[2]
      trunk GEN inside NOM hollow / empty DAT become-CONT
      'The tree trunk has become **hollow/empty** inside.'
  b. *Ōkiku naru to ki no naka ga **kara** ni narimasu.*[3]
      big   become when tree GEN inside NOM empty DAT become
      'When the tree grows it becomes hollow (lit. '**empty**') inside.'

(5) Russian
  a. *Šar   i **polyj**   cilindr <…>*
      sphere:NOM.SG and hollow:NOM.M.SG cylinder:NOM.SG
      podvešeny   k koncam rychaga   i
      hang:PTCP.PASS.PL to end:DAT.PL lever:GEN.SG and
      pogruženy   v vodu.   (RNC)
      immerse: PTCP.PASS.PL in water:ACC.SG
      'A sphere and a **hollow** cylinder <…> are attached to the ends of a lever
      and immersed in water.'
  b. *Dlja etogo   nužno vzjat' nebol'šoj   **pustoj***
      for this:GEN.N.SG need take:INF smallish:ACC.M.SG empty:ACC.M.SG
      cilindr –   lučše vsego   metallišeskij   ili
      cylinder.ACC.SG better all:GEN.N.SG metallic:ACC.M.SG or
      derevjannyj.   (RNC)
      wooden:ACC.M.SG
      'For this purpose we need a small hollow (lit. '**empty**') cylinder – one made
      of wood or metal would be best.'

Korean is notable for having two terms for EMPTY, which, however, stand in a dif-
ferent opposition. There are two near-synonyms, *pita* and *thengpita*, where *theng-
pita* is derived from *pita*. The principle behind their distribution seems to be that
thengpita describes deep containers, while *pita* describes shallow containers, and
also appears in a variety of other contexts, so that the opposition might rather be
generalized as 'deep, enclosed containers' vs. 'other containers', as illustrated in (6).

(6) Korean
  a. *Insayng-un **thengpin** pyeng-ul kathta.*   (Sejong)
      life-TOP   empty   bottle-ACC resemble
      'Life is like an **empty** bottle.'
  b. *Ta mandun kes-un **pin** cayngpan-ey nohatwunta.* (Sejong)
      all make   thing-TOP empty tray-LOC   put
      'All the ingredients are placed on an **empty** tray.'

---

2. <http://nature-makino.sakura.ne.jp/yama.html>

3. <http://www.tees.ne.jp/~takigasira/hana.html>

Situations that would fall under 'hollow' are covered by *thengpita* 'empty and deep':

(7) Korean
     I   300 millimithe-uy   kacwuk sinpal-un yongkwuk-uy namsep-wueyse
     this 300 millimeter-GEN leather boot-TOP England-GEN Southwest-LOC
     thengpin** namwu cwulkie-yse **palkyento-yessta.   (Sejong)
     empty   tree   trunk-LOC be.found-PST
     'This 300-millimeter long leather boot was found inside a **hollow** tree trunk
     in the South-West of England.'

Generally speaking, semantically the terms which describe hollow shape are
one-place predicates. If the concepts of container and contents are to be defined
via their relation to each other, the terms for 'empty of contents' will be two-place
predicates by definition. The syntactic expression of the contents varies; thus, on the
one hand, our sample has English, Spanish, and Japanese in which the absent con-
tents can be expressed as a subject (8a) or as a complement (8b, c) of the predicate.

(8) a. Japanese
         *<u>Kōhī</u> ga **kara** ni   nat-tara   sugoi hayasa de sosoide-kureru.*
         coffee NOM empty DAT become-CVB great speed INS pour-BEN
         'If you drink up your coffee (lit. 'if your <u>coffee</u> becomes **empty**'), she pours
         it really quickly.'   (BCCWJ)
     b. English
         *The bus was **empty** of any <u>children</u>.*
     c. Spanish
         *Lo-s   edificio-s   estaban   **vací-o-s**   de <u>personal</u>, <u>equipo</u>*
         DEF.M-PL building-PL be:IPF.3PL empty-M-PL of personnel equipment
         y   <u>mobiliario</u>.   (CE)
         and furniture
         'The buildings were **empty** of <u>staff, equipment and furniture</u>.'

On the other hand, the Serbian adjective *pust* and Russian *pustoj* take only one sur-
face argument – the container. In (9a) the prepositional phrase *bez mebeli* 'without
furniture' is not an argument of *pustoj*. The phrase modifies the noun *komnata*
'room' and could be used with adjectival modifiers other than *pustoj*, e.g. *malen'kij*
'small'.

(9) Russian
     a. *v malen'koj / **pustoj**   komnate   bez   mebeli*
         in small:LOC.F.SG /empty:LOC.F.SG room:LOC.SG without furniture:GEN
         'in a small / **empty** room without furniture'   (RNC)

b. *v **pustoj**   *mebeli / ot mebeli / mebel'ju*
   in empty:LOC.F.SG furniture:GEN / from furniture:GEN / furniture:INS
   komnate
   room:LOC.SG
   *'in a room **empty** of/from/by furniture'

The situation where one of an adjective's semantic arguments cannot be realized as its syntactic argument is not unique to the EMPTY domain. Consider, for example, the English adjective *dry*. The situation it describes has elaborate structure which can even denote the type of the liquid – a dry well has no water, a dry twig no sap, etc. – but the implicit arguments 'water' and 'sap' cannot be expressed as complements to *dry*.

It may also be possible that metaphoric contents are expressed more frequently than physical ones. In Spanish and English, metaphoric examples similar to (10) are more numerous than examples with physical arguments such as (8b) and (8c).

(10) Spanish
*Sin   recuerdo, la   esperanza est-ar-ía   **vací-a**   de*
without memory DEF.F.SG hope   be-INF-IRR.3SG empty-F.SG of
*significado, y   por sobre tod-o,   **vací-a   de gratitude***   (CE)
meaning   and for above all-M.SG empty-F.SG GEN gratitude
'Without memory, hope would be **empty** of meaning, and above all, **empty** of gratitude.'

Thus, the points where typological variation occurs in the domain of EMPTINESS are (1) whether 'hollow' is always expressed by a specific term (alternatively, the term for 'empty of contents' can occur in the same contexts as 'hollow', or there can be no separate term for 'hollow' whatsoever); (2) whether contents can be expressed as an argument of the term for 'empty of contents'; (3) whether shallow and deep containers are lexically opposed.

2.2   Adjacent domains

There are several semantic domains which typically overlap with that of EMPTINESS. Their vocabulary can either compete with the terms for 'empty', or describe situations from a different perspective, which makes it necessary to analyze them here. These domains are BARENESS, CLEANNESS, and POSSESSION.

2.2.1   'Bare'
'Bare' is also one of the two-place predicates of qualities, and its semantic makeup is very similar to that of 'empty of contents'. Still, 'bare' and 'empty' differ in that 'bare' focuses on an object's surface and 'empty' on what is inside the object, but

both of them mean that some expected component of the situation is absent. In the case of an empty container, the contents are expected but absent. 'Bare', similarly, describes absence of an expected covering for an object.

It is well known that languages can conceptualize flat surfaces as containers even when the former do not physically enclose their contents (cf. Herskovits 1986; Vandeloise 1991, 1994). In this regard it is quite natural for the domains of BARE-NESS and EMPTINESS to overlap when describing such container-surfaces.

Thus sometimes a space, physically the same, can be described as either empty or bare (or *blank*, a special semi-metaphorical case of emptiness; see below). It is beyond our present task to determine what containers can be characterized as 'bare', that is, the ones that have their contents conceptualized as a covering. One clear case is terrain, which can be seen either as physically 'covered' with vegetation or buildings or as functionally 'containing' them. In the latter case lexemes for 'empty of contents' can be used to describe the absence of these functional elements, as in (11).

(11) Khanty[4]
      *tătən   isa **tăλ**   χar taχi, neməλt ăn enm-əλ.*
      here.LOC all empty field place nothing NEG grow-NPST[3SG]
      'This field is quite **empty**, nothing is growing here.'

Our sample also has several instances of what at first glance appears to be a special term for 'empty' restricted to terrain and large spaces such as rooms. On closer examination, though, the specific nature of the 'container' is not the only, and even not the most prominent, semantic component of these terms.

One such word is the Russian adjective *pustynnyj* lit. 'wilderness-like' (*pustynja*, the word for wilderness or desert, is in its turn derived from *pustoj* 'empty of contents'). The term is used to describe wide, usually open, spaces, such as shores or highways, but sometimes also large rooms without people:

(12) Russian
      *My xodili   po   xolodnym, **pustynnym**   zalam*
      we walk:PST.PL through cold:DAT.PL deserted:DAT.PL hall:DAT.PL
      muzeev   zapadnoj   živopisi.   (RNC)
      museum:GEN.PL western:GEN.F.SG painting:GEN
      'We wandered through cold, **empty** halls of the museums of Western art.'

What is important about this absence of people, however, is that it is supposed to invoke an emotion, usually of loneliness, as imagining oneself in an actual wilderness.

---

4. All Khanty examples are from the fieldwork materials collected in the linguistic expedition to Tegi, Khanty-Mansi Autonomous Territory, 2010, by N.A. Murav'ev.

Pustynnyj is a rather literary, stylistically marked word, and is used rather to convey emotion than to describe the physical characteristics of a place.

The Korean term *konghehata*, derived from the Chinese loan-word *konghe* 'emptiness', is in many respects similar to *pustynnyj* – it also tends towards a more formal style and can have a connotation of sadness. Thus in (13), where *konghehata* is used together with *thengpita* 'empty and deep', the combination is not redundant because *thengpita* supplies the physical, and *konghehata* the emotional, aspect of the scene.

(13)   Korean
Nay palsoli-nun **thengpin** *kenmwul-ey* **kongheha-key** *wullye phec-yessta.*
my footstep-TOP empty   building-LOC empty-ADV   ring   spread-PST
'In the **empty** building my footsteps echoed **desolately**.'   (Sejong)

Mandarin has the compound *kōngkuàng* (lit. 'empty-waste') 'deserted', another rather literary term whose usage and connotations are similar to those of the Russian and the Korean words discussed above. *Kōngkuàng* exclusively describes broad flat spaces with no plants, buildings, or people, as well as rooms with no people and furniture, often implying a forlorn, neglected feeling:

(14)   Mandarin
Tā zài **kōngkuàng** *de   hǎitān shàng néng kàndào yī   gè   gūdú   de*
she in empty   ATR beach on   can see   one CLF lonely ATR
shēnyǐng.   (CCL)
shadow
'She could see a lone figure on the **deserted** beach.'

Terms for 'empty of contents', in their turn, are known to cross over into the domain of bareness, yet only sporadically. One of the meanings of the Japanese compound *karaashi* lit. 'empty feet' is 'barefoot', and a similar expression for 'bareheaded', *pus-togolovyj* lit. 'empty-headed' or *pustovolosyj* 'empty-haired', exists in some Russian dialects according to (Tolstaya 2008: 90). It is interesting that *pustovolosyj* literally means 'with uncovered hair' and not 'hairless' ('empty of hair'); apparently such compounds are formed with the name of the container rather than the contents (more examples of compounds are discussed in Section 2.3).

In Khanty the incursion into 'bare' occurs not only from *tǎł* 'empty', but also from one of the adjectives for 'full', *tełǎŋ*. *Tǎł* 'empty' can mean 'bare', and *tełǎŋ* 'full' can mean 'covered', for example, *tǎł/tełǎŋ mejvł* 'bare (lit. 'empty') / clothed (lit. 'full') torso'. Another Khanty adjective of fullness, *tełjeva,* in the same context would mean that the chest is covered with a large number of badges or other decorations.

2.2.2   'Clean'

'Clean' brings to our attention another kind of non-prototypical container, midway between the physical and the metaphorical senses of 'empty of contents' – namely, 'media as containers for information.'

We may define the basic meaning of 'clean' as 'free from dirt', with 'dirt' roughly meaning 'a substance that should not be on or within the object'. Marks on the object's surface can be viewed as a special case of 'dirtiness'.

In the case of media, marks (text, drawing, recording, etc.) are the physical manifestation of contents. Contents are not normally described as making the media 'dirty', but their absence equals the absence of surface marks, which allows the terms for 'clean' to be used:

(15) Spanish
 toalla / página **limpi-a**
 towel / page   clean-F.SG
 'a **clean** towel / page'

In this situation 'clean' becomes almost synonymous with 'empty of contents'. In Russian, the choice between *pustoj* 'empty of contents' and *čistyj* 'clean' depends on whether the marks are expected as a part of the medium's normal state. Thus, blank pages in a book would be described as 'empty', rather than 'clean', because books are expected to have text. Unmarked leaves in a notebook, on the other hand, are 'clean', because this is the state in which they are capable of their normal functioning:

(16) Russian
 Perelistyvaju tetrad'   *do* **čistyx**   *stranic.*   (RNC)
 leaf:PRS.1SG notebook:ACC.SG to clean:GEN.PL page:GEN.PL
 'I am leafing through the notebook until I come to the **clean** pages.'

This usage of 'clean' does not have 'dirty' for its antonym either in Russian or in Spanish.

Empty media as containers may have a dedicated lexeme, like English *blank*, or they may fall under the general term for 'empty of contents', as in Japanese, Mandarin, and Serbian in (17).

(17) a. Japanese
 Kara *no CD o   ire-ta   toki wa …*   (BCCWJ)
 empty GEN CD ACC insert-PST time TOP
 'When you insert a blank (lit. '**empty**') CD…'
 b. Mandarin
 Qǐngjiào, dǎyìnjī wúgù   zǒu **kōng** *zhǐ, zěnme bàn*   (CCL)
 consult   printer no.reason go empty paper how   do
 'Could you tell me what to do when a printer for no reason prints **empty** pages?'

    c.   Serbian
       Cela   *traka*   *je*   ***prazna.***   (Glosbe_SR)
       entire:NOM.F.SG tape:NOM.SG be:PRS3SG empty:NOM.F.SG
       'The entire tape is blank (lit. **empty**).'

Korean uses *pita*, the term for shallow containers, rather than *thengpita* which is designated for deep containers:

  (18) Korean
       Ku-nun caki   *kongchayk-eyse **pin /*thengpin** pheyci han cang-ul*
       he-TOP oneself notebook-ABL empty / empty page   one sheet-ACC
       ttuten-ayssta.   (Sejong)
       take.off-PST
       'He tore an **empty** / *'**empty and deep**' page out of his notebook.'

2.2.3   *Possession*

Possession is a situation that can be conceptualized in terms of location (Taylor 1989; Heine 1997). Since contents are located in a container, it creates the possibility for containers to be described in terms of possession – a full container is 'occupied', and an empty one is 'free':

  (19) a.   Spanish
          el   *únic-o*   *asiento **libre/vací-o***
          DEF.M.SG only-M.SG seat   free /empty-M.SG
          'the only **empty/free** seat'
      b.   Russian
          Najdi   ***svobodnoe/pustoe***   *kreslo.*
          find:IMP.SG free:ACC.N.SG / empty:ACC.N.SG seat:ACC.SG
          'Find a **free/empty** seat.'

In Japanese both meanings – 'empty of contents' and 'free, unoccupied' – are covered by the verb *aku*, the main sense of which is '(intransitive) open'. It is typically used in the continuous or perfect form, and takes both humans and objects as contents, as shown in (20).

  (20) Japanese
      a.   *Iriguchi doa ga **ai-ta**   mama no jōtai de…*   (BCCWJ)
          entrance door NOM open-PST as   GEN state be.CVB
          'The door at the entrance was still **open**'.
      b.   ***Ai-ta***   *hako o*   *yōi*   *shimasu.*   (BCCWJ)
          open-PST box ACC preparation do
          'We prepare an **empty** (lit. "open") box.'

c. ***Ai-ta***   *seki*   *o*   *sagashite-iru*   (BCCWJ)
open-PST seat ACC search-CONT
'He is looking for an **empty** (lit. "open") seat.'

Normally the physical presence of contents is necessary for a container to be considered occupied. People are a notable exception – they can exercise symbolic presence, and possession, without being actually present in the location at the moment. In fact, in languages such as Russian the terms for 'empty' are ambiguous when they describe locations. Contexts such as *pustoj dom* 'an empty house' can receive a possessive interpretation: 'an ownerless, uninhabited house'; for spaces with a different functionality, *pustoj* can mean the situational physical absence of people (but not necessarily of other contents: the 'empty museum halls' in (21a) are not literally empty, because the exhibits are present). Finally, in (21b) it is the inanimate contents, the exhibits, which are absent:

(21)   Russian
     a.   *tol'ko za*   *odnu*   *nespešnuju*   *progulku*   *po*
         only   during one:ACC.F.SG slow:ACC.F.SG walk:ACC.SG through
         pustym   *zalam*   *muzeja*   (RNC)
         empty:DAT.PL hall:DAT.PL museum:GEN.SG
         'during a single leisurely stroll through the **empty** halls of the museum'
     b.   ***Pustymi***   *ostalis'*   *ermitažnye*   *zaly* –
         empty:INS.PL be.left:PST.PL Hermitage:NOM.PL hall:NOM.PL
         bol'šuju   *čast'*   *eksponatov* *vyvezli*   *v*
         bigger:ACC.F.SG part:ACC.SG item:GEN.PL evacuate:PST.PL to
         Sverdlovsk.   (RNC)
         Sverdlovsk:ACC
         'The halls of the Hermitage Museum were left **empty** – most of the collection was evacuated to Sverdlovsk.'

A standard dictionary of Mandarin Chinese[5] specifically mentions a similar ambiguity of the compound ***kōngfáng*** 'empty house', defining its meaning as 'a house where no things are stored or no people live'.

     The ambiguity between the absence of people and objects is not characteristic of all languages. Korean distinguishes the two situations: *pita*, the general term for 'empty of contents', is used for the absence of human occupation, and *thengpita* 'empty and deep' for the absence of objects, as illustrated in (22).

---

5.   *Xiàndài Hànyǔ cídiǎn* (Dictionary of Modern Chinese). 6-th edition. Běijīng: Shāngwù yìnshū guǎn 2012.

(22)   Korean
    a.   *Sakwaha-myen **pin** kayksil-i eps-tako hayss-eyoo.* (Sejong)
       apologize-CVB empty room-NOM not.exist-QUOT say-PST
       'I apologized and said there were no **vacant** rooms.'
    b.   *Kakwuka epse pang-ani **thengpin** kes kathta.* (Sejong)
       furniture-NOM not.exist room-NOM empty thing resemble
       'Without furniture the room looks **empty**.'

In Serbian the adjective *pust* historically used to mean both the situational absence of people and the lack of a permanent occupant, but has narrowed its scope to mean primarily the latter. The concept of situational absence has been taken over by the main term for 'empty of contents', *prazan*.

(23)   Serbian
    a.   *dodatna **prazna** soba* (Glosbe_SR)
       additional:NOM.F.SG empty:NOM.F.SG room:NOM.SG
       'an additional **empty** room'
    b.   *na **pustom** ostrvu* (Glosbe_SR)
       on empty:LOC.N.SG island:LOC.SG
       'on a **desert** island'

We have seen that the domains of BARENESS, CLEANNESS and POSSESSION show a typological tendency to overlap with EMPTINESS. Another notable intersection of domains can be seen in Korean. Korean has a usage where *pita* 'shallow and empty of contents' crosses into the domain of INCOMPLETENESS:

(24)   Korean
    a.   *sipman wen-eyse chen wen-i **pinta*** (Sejong)
       10 000 won-LOC 1 000 won-NOM empty
       'one thousand won short (lit. '**empty**') of ten thousand'
    b.   *Mwulken kayswu-ka **pica** sacang-un congepwen-ul*
       thing number-NOM empty.CVB boss-TOP worker-ACC
       uysimh-ayssta. (Sejong)
       doubt-PST
       'After some things were missed (lit. '**empty**'), the boss started suspecting his staff'.

Only Korean in our sample manifested this connection between EMPTINESS and IN-COMPLETENESS. The domain of FULLNESS, however, has a strong connection to COM-PLETENESS: terms for 'full' mean 'complete' in several languages (see Section 3.2).

2.3 Semantic extensions

In this section we shall go over some cases where the terms in the domain of EMPTI-NESS are used in non-physical sense. Some of these usages can be seen as metaphors, conceptualizing various abstract entities as containers and contents, such as TIME IS A CONTAINER FOR EVENTS, TEXT IS A CONTAINER FOR INFORMATION, HEAD AND EYES ARE CONTAINERS FOR INTELLECTUAL AND EMOTIONAL ACTIVITY (cf. Lakoff & Johnson 1980; Grady 2007). Other usages appear too vague and might be better described as terms undergoing semantic bleaching and losing more or less of their precise meaning. Even the most abstract usages, however, share one semantic component which can be broadly labeled "betrayed expectations": an expected component of the situation is absent, just as contents are absent form an empty container.

S. Tolstaya notes that the absence of contents for 'empty' is seen as an anomaly, and that adjectives which describe some deficiency as abnormal invariably develop negative connotations (Tolstaya 2008: 85, 93). Her observation is made for Slavic languages, but the rest of our sample appears to support it.

Several other domains discussed in this volume, besides EMPTY, contain adjectives with evaluative connotations, positive or negative (see Chapters 6 on SMOOTH and 2 on SHARP). Usually it is the underlying situation of functional interaction with objects which determines whether a term receives a positive or negative connotation. This can be seen as supporting A.Wierzbicka's view that language treats qualities as inherently anthropocentric (Goddard & Wierzbicka 2007). An empty container is "bad" insofar as it is not fulfilling its function of holding useful contents.

Descriptions of time as 'empty of contents' vs. 'free' are illustrative in this respect. The contents of time as container are events and activities which normally 'fill' or 'occupy' time. Describing time as 'empty of contents' in (25) and (26a) carries a negative connotation. When the absence of these events and activities is perceived as a good thing, terms for availability can be used, as in (26b).

(25) Spanish
*Genial, un-a   hora **vací-a**   con es-e   idiota con   el*
brilliant INDF-F.SG hour empty-F.SG with that-M.SG idiot   with DEF.M.SG
que no ten-ía   nada   que habl-ar.   (esTenTen18)
REL NEG have-IPF.1SG nothing REL talk-INF
'Great, an **empty** hour with that idiot with whom we don't even have anything to talk about.'

(26) Japanese
a. *Kanojo wa   tatta hitori de, **utsuro** na   hibi o   sugoshite-i-ta*
she   TOP only one   INS hollow ATR days ACC spend-CONT-PST
'She was spending empty (lit. **hollow**) days all by herself.'   (Kenkyusha)

    b.   *Chotto jikan ga* **ai-ta** *toki ni kaimono suru toki datte aru.*
        briefly time NOM open-PST time DAT shopping do   time also be
        'Or sometimes you do some shopping when you have a little free time (lit.
        "when time has **opened** a little")'.   (BCCWJ)

TIME IS A CONTAINER is one of the ontological metaphors discussed in (Lakoff &
Johnson 1980: 60). Other metaphors mentioned in that work which are relevant to
the domain of EMPTINESS include LINGUISTIC EXPRESSIONS ARE CONTAINERS FOR
MEANINGS (presented as a part of the broader metaphor of language as conduit,
ibid: 10, 12) and THE EYES ARE CONTAINERS FOR THE EMOTIONS (ibid: 51). In most
of our examples where terms for emptiness are applied to words, texts and eyes,
they carry negative connotations.

(27) Spanish
    a.   *un*   *discurso retóric-o*   *y*   **huec-o**   (CREA)
       INDF.M.SG speech rhetorical-M.SG and hollow-M.SG
       'a speech made of empty (lit. '**hollow**') rhetoric'
    b.   *un*   *discurso* **vací-o**   *de contenido*   (CE)
       INDF.M.SG speech empty-M.SG of content
       'a speech **devoid** of any content'

(28) Khanty
    naŋ i   *tum jasəŋ jast-əs-ən,*   *pa jasăŋ-λ-ăn* **tăλ**   *jasăŋ-ət.*
    you one that word say-PST-2SG ADD word-PL-2SG empty word-PL
    'You've said everything with one word, all the other words are **empty** (= redundant)'.

(29) Mandarin
    Zhè piān lùnwén hěn **kōng** *méi yǒu nèiróng.*   (CCL)
    this CLF article very empty not have content
    'The article is quite **empty**, there is nothing in it.'

We repeatedly find eyes described as 'empty' or 'hollow' to indicate a state of stupor,
as in (30):

(30) a. Spanish
       un-a   *mirada* **vací-a**   (Glosbe_ES)
       INDF-F.SG look   empty-F.SG
       'a blank (lit. '**empty**') look'
    b. Khanty
       tăλ   *sem-ən want-əλ,*   *moλti*   *nuoməs-man.*
       empty eye-LOC watch-NPST what.INDF think-CVB
       'He's staring with **empty** eyes, upset about something.'

   c.   Japanese
       Karada wa yasehosori, me mo **utsuro** *de at-ta.*   (Kenkyusha)
       body   TOP macerate eye too empty INS be-PST
       'His body was emaciated and his eyes were blank (lit. '**empty**').'
   d.   Russian
       Im   eto   uže   neinteresno. **Pustye**
       they.DAT this.NOM:N.SG already uninteresting empty:NOM.PL
       glaza.   (RNC)
       eye:NOM.PL
       'That already doesn't interest them. Their eyes are **empty**.'
   e.   Korean
       Ku-nun yenkuk-i   kkuthna-ko mwutay-lul **konghehan** *nwun-ulo*
       he-TOP play-NOM end-CVB   stage-ACC empty   eye-INS
       palapo-assta.   (Sejong)
       stare-PST
       'The play was over, but he kept staring at the stage with **empty** eyes.'

This metaphor has a physiological basis: the level of alertness affects the appearance of the eyes, so that awareness is read "in" them and absent "from" them; but we could find no corresponding metaphor for 'full.'

    Another body part which functions as a metaphorical container is the head. As the location of the brain, it is widely believed to metaphorically contain the intellect – or the lack of it, with the expected negative connotation. Serbian has *prazna glava* lit. 'empty head' and *šuplja glava* lit. 'hollow head', both meaning a stubborn and stupid or absent-minded person. See also the examples from Korean and Mandarin in (31) (note that in Korean the term for 'empty and deep' containers that also covers 'hollow' is used).

(31) a.   Korean
       Kunye-nun kakkumssik meli-ka   **thengpin** *salam-chelem*
       she-TOP   episodic   head-NOM empty   person-like
       hayngtong-ul hay.   (Sejong)
       behavior-ACC do
       'Sometimes she behaves like an **empty**-headed person'.
    b.   Mandarin
       Nǎodai **kōng kōng** *de, shénme dōu xiǎng   bù qǐlái.*   (CCL)
       head   empty empty ATR what   all   remember not can
       'My head is quite **empty**, I can't remember anything'.

The Russian dialectal usage of 'empty-headed' to mean 'bareheaded' was mentioned in connection with 'bare'. Example (32) is a Russian army saying which means that, according to the regulations, a soldier's head must be covered when giving a military salute:

(32)   Russian
    *K **pustoj**   golove   ruku   ne prikladyvajut.*
    to empty:DAT.F.SG head:DAT.SG hand:ACC.SG not touch:PRS.3PL
    'No touching your hand to a bare (lit. **empty**) head.'

This expression might have originated in a dialect, but is currently used as metaphor, implying that the soldier who salutes bareheaded has no brain.

    'Empty hands' is an expression found in most of our sample. The non-physical usages seem to be metonymic rather than metaphoric. Hands are a natural container for whatever the situation requires – presents when visiting, trophies when returning from a hunt, weapons, etc., – so the image of not carrying something physically can stand for a broader meaning of a relevant component absent from the situation. Languages may differ slightly in the preferred contexts (in Russian, for example, the most common situation for 'hands as an empty container' is having no presents to give or not having received any, cf. (33a)), but the logic is basically similar in other languages.

(33)   a.   Russian
    *K činovnikam s   **pustymi**   rukami   ne xodjat.*
    to official:DAT.PL with empty:INS.PL hand:INS.PL not go:PRS.3PL
    'One doesn't visit government officials **empty**-handed (without a bribe).'
                         (RNC)

    b.   Korean
    ***Pin** son-ulo kako siph-ci-nun   anha.*   (Sejong)
    empty hand-INS go.CVB want-NEG-TOP NEG
    'I don't want to go **empty**-handed (without a present).'

    c.   Serbian
    *Ne mogu   opet da odem   kući **praznih***
    not can:PRS.1SG again CONJ go:PRS.1SG home empty:GEN.F.PL
    ruku.   (Glosbe_SR)
    hand:GEN.PL
    'I can't go home **empty**-handed again (having failed to get anything).'

    d.   Spanish
    *Napoleón se   qued-aba   con la-s   mano-s **vací-a-s**.*
    Napoleon REFL leave-IPF.3SG with DEF:F-PL hand-PL empty-F-PL
    'Napoleon was left **empty**-handed (without trophies).'   (CREA)

    e.   Khanty
    *χoλijewa noχ aλemə-s-aj-ət,   **tăλ** jɔš-ən   χaś-s-əm.*
    everything up take-PST-PASS-3PL empty hand-LOC be.left-PST-1SG
    'They took everything, I was left **empty**-handed (without possessions).'

The Mandarin compound *kōngshǒu* 'empty-handed' can mean 'unarmed'. In Japanese, a similar compound *karate* 'empty hands' has provided the name for a style of fighting without weapons; when the verb *aku* 'be open, empty or available' is used instead, the expression moves away from the physical sense to mean 'not busy', as in (34).

(34) Japanese
    *Te   ga   **aite**-iru   node   tetsudaimasu.*
    hand NOM open-CONT because help
    'I'm free (lit. 'hands are **open**), so I'll help you.'

It is worth mentioning that Russian and English express the absence of weapons or instruments with a term from an adjacent domain, ***bare**handed*:

(35) Russian
    *Tebja   možno s   **golymi**   rukami   na tanki   posylat'!*
    you.ACC can   with bare:INS.PL hand:INS.PL on tank:ACC.PL send:INF
    'One could send you to fight tanks with **bare** hands!'   (RNC)

Examples (33–35) show that the combinations of 'hands' with terms for 'empty of contents' or terms from adjacent domains are often lexicalized and develop unpredictable nuances of meaning which vary across languages. The same tendency towards lexicalization can be seen in other collocations. The terms for emptiness simply state that something is missing, the nature of the missing component is determined by convention and/or deduced from the situation, requiring extralinguistic knowledge.

Japanese has a number of compounds with *kara* 'empty of contents' as the first component, the best known of which is *kara**oke*** (lit. 'empty orchestra') which refers to instrumental accompaniment without the vocals. Another one, *kara**cha*** (lit. 'empty tea') 'tea without sweets', has a parallel in the somewhat dated Russian *pustoj čaj* (lit. 'empty tea') with the same meaning, *pustye šči* 'cabbage soup without meat' (lit. 'empty cabbage soup') and, with a term from the adjacent domain of BARENESS, *golyj ris* 'rice without meat or sauce' (lit. 'bare rice').

"Betrayed expectations" expressed by the terms of emptiness are not limited to physical objects. Many examples deal with highly abstract concepts such as the expected relation between speech and action, action and results, or behavior and its cause.

(36) a. Russian
Pustoe   *zanjatie –*   *rasskazyvat' kartinu.*
empty:NOM.N.SG business:NOM.SG tell:INF   painting:ACC.SG
S"jezdi   posmotri.   (RNC)
go:IMP.SG look:IMP.SG
'It's pointless (lit. '**empty** business') trying to describe a painting. Go have a look at it.'

    b. Serbian
On nije   doživljavao   svoju   žrtvu
he   be:NEG.PRS.3SG feel:PST.PTCP.M.SG his:ACC.F.SG sacrifice:ACC.SG
*kao sujetnu   ili **praznu**.*   (Glosbe_SR)
as   vain:ACC.F.SG or empty:ACC.F.SG
'He did not feel his sacrifice to be a vain or **empty** one.'

    c. Spanish
*la-s   ya   familiar-es mentira-s y   promesa-s **vací-a-s***
DEF.F-PL already familiar-PL lie-PL   and promise-PL empty-F-PL
'the already familiar lies and **empty** promises'   (CE)

(37) a. Serbian
*Ne padam   ja na tvoje   **prazne***   *pretnje.*
not fall:PRS.1SG I   on your:ACC.F.PL empty:ACC.F.PL threat:ACC.PL
'I don't buy your **hollow** threats.'   (Glosbe_SR)

    b. *Mandarin*
***kōng**máng*
'empty (=pointless) efforts'

    c. Japanese
***kara**ibari*
empty.boasting
'bravado'

    d. Spanish
*El   niño se   puso   muy **huec-o**   cuando lo*
DEF.M.SG boy   REFL put.AOR.3SG very hollow-M.SG when   him
nombr-aron   ganador.   (Collins)
name-AOR.3PL winner
'The boy was very conceited (lit. '**hollow**') when he was declared the winner.'

An interesting example of further desemantization and semantic bleaching was found in Khanty. The adjective *tăł* 'empty' can also mean 'only', losing most of its physical meaning. Native speakers point out that 'empty paper' in (38) cannot be interpreted as 'clean paper with no writing':

(38) Khanty
      manema mij-e   **tăλ**   *nepek*
      I.DAT   give-IMP.SG empty paper
      'Give me just the paper, nothing else' (lit. 'Give me **empty** paper').

Such development, from the situation of physical absence to a highly abstract meaning, is typical of terms of lack or absence in general (cf. Tolstaya 2008). Good examples are the English adjective *pure* 'without admixtures' in such contexts as *out of pure spite* or Russian *golyj* 'bare' in *golyj entuziazm* 'sheer enthusiasm'.

3. FULL

We now turn to the semantic domain of FULL. It has been noted that for terms that express qualities, an antonym is almost never a simple negation. Goddard & Wierzbicka (2007) shows in detail the differences between the antonymic pairs *rough* and *smooth, hard* and *soft,* and some other pairs, and the antonyms *sharp* versus *blunt* and *rough* versus *smooth* are discussed in this volume. Similarly, 'full' is not a straightforward opposite of 'empty'. The whole structure into which 'full' develops from its core situation is different. The next sections cover the physical sense of 'full' and its syntax, its relation to some adjacent semantic domains, and its main semantic shifts.

3.1 The main (physical) sense

3.1.1 *Container and quantity*
One respect in which 'full' and 'empty' domains are similar is that both domains involve containers. If an empty container lacks contents, a full container holds the maximum possible amount of them, i.e. container X is *full* of Y if so much Y is located in X that no further amount can be placed there, as in (39a–e).

(39) a. Mandarin
          *Tǒng   yào   **mǎn** le.*   (CCL)
          bucket almost full   MOD
          'The bucket is almost **full**.'
      b. Serbian
          *Vreće   i   koferi   **puni**   gotovine*
          sack:NOM.PL and suitcase:NOM.PL full:NOM.M.PL cash:GEN.SG
          'Sacks and suitcases **full** of cash'   (Glosbe_SR)

c. Russian
*s   tarelkoj,   **polnoj**   ogurcov,   pomidorov,*
with plate:INS.SG full:INS.F.SG cucumber:GEN.PL tomato:GEN.PL
zelenogo   luka,   salata.   (RNC)
green:GEN.M.SG onion:GEN.SG lettuce:GEN.SG
'With a plate **full** of cucumbers, tomatoes, green onion and lettuce'

d. Khanty
*χojat   λoŋ-əs   **teλijewa/teλ/teλeŋ** padnɔs piλ-ən.*
who.INDF enter-PST full   tray   with-LOC
'A person with a **full** tray came in.'

e. Spanish
*un-a   caja **llen-a**   de libro-s   de poesia.*   (CE)
INDF-F.SG box full-F.SG of book-PL of poetry
'a box **full** of poetry books'

The name of the container may undergo a metonymic shift and signify a quantity of contents rather than the container itself as a physical object. This shift is usually accompanied by some restrictions on the syntactic position of 'full'. Compare the following examples where 'full' can only be used in preposition to the head noun in the quantitative sense:

(40) English
a. *He drank a **full** <u>glass</u>* (container as quantity) *of water* (=contents).
b. **He drank a <u>glass</u> **full** of water.*

(41) Russian
a. *On   vypil   **polnyj**   <u>stakan</u>   vody.*
he.NOM drink:PST.M.SG full:ACC.M.SG glass:ACC.SG water:GEN.SG
'He drank a **full** glass of water' (adjective *polnyj* 'full' modifies *stakan* 'glass' as an attribute).
b. **On   vypil   <u>stakan,</u>   **polnyj**   vody.*
he.NOM drink:PST.M.SG glass:ACC.SG full:ACC.M.SG water:GEN.SG
'He drank a glass that was **full** of water' (adjective *polnyj* 'full' is the predicate of the relative clause which modifies *stakan* 'glass').

3.1.2   'Full' vs. 'not hollow'

The terms for 'empty of contents' can often be used for hollow shape. The terms for 'full', unlike them, usually do not describe a shape. In Spanish it is possible to speak of a *nuez **llena*** lit. 'a full nut', as opposed to a hollow one, but nuts are cultivated specifically for their kernels, and it can be said that in this respect the shell of a nut is similar to a manufactured container. The only language in our data that appears to use 'full' as an opposite of 'hollow' is Serbian, where *pun* 'full' can be seen in contexts such as (42):

(42) Serbian

*To   je   **puni**   kamen.*   (Glosbe_SR)
this:NOM.N.SG be:PRS.3SG full:NOM.M.SG stone:NOM.SG
'It's solid (lit. '**full**') stone'.

Alternatively, *čvrst* 'solid, hard' is used, e.g. *čvrsta stena* 'solid rock.'

The opposite of 'hollow' is mostly described by terms from several semantic domains adjacent to FULL, namely SOLID, INTACT, and COMPLETE. Specific terms for the antonym of 'hollow' may also exist. In Mandarin the opposite of *kōngxīn* 'hollow' lit. 'empty-centered' is the compound *shíxīn* lit. 'solid-centered', e.g. *shíxīn qiú* 'a solid sphere'. Russian has the adjective *cel'nyj* derived from *celyj* 'whole':

(43) Russian

*Ego **cel'nyj**   korpus   ne progoraet   ot   vysokix*
it.GEN solid:NOM.M.SG frame:NOM.SG not burn:PRS.3SG from high:GEN.PL
temperatur.   (RNC)
temperature:GEN.PL
'Its [= a heater's] **solid** frame is resistant to high temperatures.'

Both the Russian and the Mandarin terms are narrow and rather technical, without any prominent metaphoric usage.

The difference between 'empty of contents' describing 'hollow' and 'full' describing 'not hollow' appears to lie in the fact which was pointed out in (Tolstaya 2008) for Slavic languages but may be more broadly relevant. Tolstaya claims that the terms for lack of something are semantically marked: they carry a negative connotation because the situations they describe are seen as abnormal. Tree trunks, walls and other non-container objects are normally not expected to be hollow; when they are, it warrants mention. Their being solid, on the other hand, is merely the default state. When it is remarked upon, it is usually to note the object's other qualities such as its texture or integrity.

Another crucial difference between the domains EMPTY and FULL is that a container's capacity is limited, and 'full', but not 'empty of contents', brings this limit into focus. We find terms for 'full to the limit', which are opposed to the main terms of fullness. Korean and Spanish have attributives, while Japanese and English use verbs which mean 'to overflow'. The opposition can be described as "technically vs. functionally full": any more contents cannot be added to an 'overflowing' container, and need not be added to a 'full' container because the optimal limit has been reached.

Liquid contents illustrate this difference well, because cups, buckets etc. in most everyday situations are filled perceptibly less than to their maximum technical capacity to avoid spilling.

(44)   Korean

    a.   *Kyewul hayssal-i   cohun nal-ey   khephi **kutukhan** can-ul*
       winter sunshine-NOM good day-LOC coffee full   cup-ACC
       tulko   hayssal-ul   culkiko siphta.   (Sejong)
       hold.CVB sunshine-ACC enjoy.CVB want
       'On a sunny winter day one wants to enjoy the sunlight with a **full** cup of coffee.'

    b.   *Mwul-i   **katukhan** yokco-ey mom-ul   tamku-myen mwul-i*
       water-NOM full   tub-LOC body-ACC immerse-CVB water-NOM
       nemchinta.   (Sejong)
       overflow
       'If you immerse a body into a bath **full** of water, the water will overflow.'

With non-liquid contents, such as laundry in (45), the terms for 'overflowing' are used as a metaphor and carry the connotation that the amount of contents is large (cf. Section 3.3.2).

(45)   Spanish
    *un-a   cesta **replet-a**   de ropa suci-a*
    INDF-F.SG basket overfull-F.SG of clothes dirty-F
    'a basket **overflowing** with dirty laundry'

(46)   Japanese
    *Tennai wa hito   de **afurete**-imashita.*   (BCCWJ)
    shop   TOP person INS overflow-CONT
    'The shop was full of ("lit. **overflowing** with") people.'

The underlying set of semantic arguments for 'full' is the same as for 'empty of contents', but for 'empty' the overt expression of the contents in some languages is limited or even impossible. In contrast, for 'full' the contents are freely expressed.

    Terms for 'full' can also modify other predicates to describe extent or degree (see Section 3.3 below). For 'empty of contents' such usage is uncharacteristic.

    In sum, FULL as a domain is structured differently than EMPTY. The former domain has no specific terms for 'not hollow', but there is a metonymic shift into the meaning of quantity and an opposition of technically vs. functionally full containers.

3.2 Adjacent domains

DIRTY and COVERED do not seem to manifest the same overlap with the domain of FULLNESS as their antonymic counterparts CLEAN and BARE have with EMPTINESS.[6] Of the domains which interact with EMPTY, only one was found to cross over into FULL: this is occupancy phrased in terms of POSSESSION. On the other hand, FULL has an overlap with COMPLETE which was found to overlap with EMPTY only in Korean, see Example (24).

3.2.1 *Possession*

Locational occupancy of a container can be expressed by terms for 'occupied', which is similar to 'unoccupied' describing empty containers.

In Section 2.2.3 we have seen that terms for 'empty of contents' describing locations can be ambiguous: they can mean current absence of people or the lack of a long-term occupant (e.g. *an empty house* can be either temporarily deserted or uninhabited). The terms for 'full' are less likely to describe occupancy. When they describe locations, the terms for 'full' for the most part can be understood straightforwardly as physical localization, i.e. that people as physical objects are present in large numbers. Such phrasing implies that no more people can get in, which is usually a hyperbole.

(47) Korean
*Hanben-un ceto keuy   namcat-ul-lo **katukhan** besu-lul thapon*
once-TOP I   almost man-PL-INS full   bus-ACC ride.PST
cek-i   iss-esseyo.   (Sejong)
time-NOM be-PST
'Once I rode on a bus **full** of men.'

In Serbian we have attested one usage of 'full' that does not mean 'currently packed to the limit':

(48) Serbian
Jedan   krevet   je   prazan,
one:NOM.M.SG bed:NOM.SG be:PRS.3SG empty:NOM.M.SG
*ostali   su   **puni**.*   (Glosbe_SR)
remaining:NOM.M.PL be:PRS.3PL full:NOM.M.PL
'One bed is empty, the others **occupied**.'

---

6.   Khanty provides a rare example of 'full' meaning 'covered', see Section 2.2.1.

A literal translation of this sentence into Russian with *polnyj* 'full' (a cognate of the Serbian adjective *pun* 'full' which is used in Example (48)) sounds odd, but if it were used, it would imply that there are several occupants to each bed. Normal locational occupancy in Russian can only be expressed by a specific term for occupancy, *zanjatyj* 'occupied'.

Apart from such examples as (48), Serbian is similar to Russian in that the main term for the presence of temporary occupants in Serbian is not *pun* 'full' but *zauzet* 'occupied':

(49) Serbian
*Nisu   svi   katovi   **zauzeti.***
be:NEG.PRS.3PL all:NOM.M.PL floor:NOM.PL occupied:NOM.M.PL
'Not all the floors are **occupied**'.   (Glosbe_SR)

3.2.2   *'Full' and 'complete'*

Completeness is a characteristic of a specific kind of objects – sets. While many objects have parts but are still seen primarily as a whole, a set is conceptualized specifically as a collection of parts. Elements of a set may form an ordered sequence, such as words in a quotation or volumes in a book series, but it is not necessary. A set or sequence is complete when no element is missing.

No elements need or can be added to a complete set, just as no contents can be added to a full container. Thus terms for 'full' can be used to describe completeness, motivated by the container metaphor: WHOLE IS A CONTAINER FOR ITS PARTS.

Another way of describing a complete set is by using the vocabulary of INTACT-NESS ('whole', 'entire'), because being undamaged implies that all parts are present.

Languages vary in describing sets and sequences as 'full', 'entire' or 'complete':

(50) a. *full name* (English), *polnoe imja* (Russian), *puno ime* (Serbian)
      b. *nombre completo* 'full (lit. 'complete') name' (Spanish), *quán míng* 'full (lit. 'entire') name' (Mandarin)

Serbian has a loan-word *kompletan* 'complete', but *pun* 'full' is also used in similar contexts:

(51) Serbian
      a. *kompletan čovečiji genom* 'the complete human genome'
      b. *puna instalacija* 'full installation' (of a program: one that includes all of the program's components)

Spanish makes a finer distinction for sets. It has two separate terms of INTACTNESS. *Completo* 'complete' is used for structured sets consisting of a definite number of more or less varying elements. *Íntegro* means 'undamaged' and is used when the

whole is continuous and more or less uniform. The two terms are illustrated in (52); both are used to describe texts, but (52a) represents a set of texts as a sequence of volumes, and (52b) – a single text as a sequence of words.

(52) Spanish
    a. *la-s*   *obra-s* ***complet-a-s*** *de lord Byron*   (CE)
        DEF.F-PL work-PL complete-F-PL of lord Byron
        '**complete** set of works of Lord Byron'
    b. *el*   *texto* ***íntegr-o***   *de-l*   *Tratado de Maastricht*   (CE)
        DEF.M.SG text whole-M.SG of-DEF.M.SG treaty   of Maastricht
        'the full (lit. **whole**) text of the Maastricht Treaty'

3.3   Semantic extensions

Semantic extensions for the terms of fullness have two major directions. A semantic shift can be based on the idea of a container as such (Section 3.3.1), giving rise to container metaphors (in this respect FULL functions similarly to EMPTY). Alternatively, metaphorization can be based on the idea of the container's maximum capacity being reached. This implies either connotations of a large quantity of contents (Section 3.3.2) or the idea of a maximum, allowing the terms for 'full' to function as intensifiers (3.3.3).

3.3.1   *Container metaphors*
Many things can be metaphorically described as containers. Among the more common metaphors in our sample are EXPRESSIONS ARE CONTAINERS FOR EMOTIONS and TIME IS A CONTAINER FOR EVENTS, which parallel the container metaphors discussed in Section 2.3.

(53) a. Mandarin
        chōngmǎn *gǎnqíng de*   *huà*
        fill   emotion ATR word
        'words **filled** with emotion'
    b. Japanese
        yume ni ***michi-ta*** *kikaku*   (Kenkyusha)
        dream DAT fill-PST   project
        'a visionary project (lit. 'a project **full** of dreams')'
    c. Spanish
        un   *año* ***plen-o*** *de actividad*   (Glosbe_ES)
        INDF.M.SG year full-M.SG of activity
        'an active year (lit. 'a year **full** of activity')'

d.   Korean
*Ney-ka   malhan taylo pyengsowa kathun himtun il-i   **katukhan***
you-NOM say.PST way   usual   same   hard   work-NOM full
nal-tul iesse.   (Sejong)
day-PL be-PST
'As you said, those were the days **full** of hard work.'

3.3.2   *Large quantity*

We have seen in Section 3.1.1 that the name of a container modified by a term for 'full' can be used metonymically to signify a quantity, rather than the container itself. The notion of the maximum limit that is present in the main sense of the terms for 'full' carries over as an evaluative connotation: the quantity is evaluated as large. This semantic component then develops into hyperbole; while there is some snow inside the boot in (54), the boot is not literally filled with snow:

(54)   Russian
*Skol'zit   i   dvaždy ostupaetsja,   nabrav   **polnyj***
slip:PRS.3SG and twice   stumble:PRS.3SG gather:CVB.PST full:ACC.M.AG
botinok   snega.   (RNC)
boot:ACC.SG snow:GEN.SG
'He slips and stumbles twice, getting one of his boots full of snow' (lit. 'gathering a **full** boot of snow').

When such hyperbolic usage undergoes a degree of semantic bleaching, the terms for 'full' develop into modifiers that simply mean 'a lot':

(55)   a.   Korean
*Saykyun-i   **katukhan** son-ulo ayin-kwa son-ul   maccapko*
bacteria-NOM full   hand-INS lover-CMT hand-ACC grab
teyithu-lul hanun cangmyen-ul sangsanghay posey-yo.   (Sejong)
date-ACC do   scene-ACC   imagine   look-IMP
'Imagine yourself on a date, holding your sweetheart's hand with a hand **full** of microbes.'
b.   Spanish
*un   pantalón de mezclilla **llen-o**   de mancha-s de pintura*
DEF.M.SG pants   of denim   full-M.SG of stain-PL of paint
'denim pants covered in (lit. '**full** of') paint stains'.   (CE)
c.   Serbian
*Skeptika   je   **puno**.*   (Glosbe_SR)
skeptic:GEN.PL be:PRS.3SG full:NOM.N.SG
'The skeptics are **numerous**.'

d.   Khanty
teλ/teλijewa soχ
full   skin
'a lot of clothes' (lit. '**full** clothes')

Russian has an adverb derived from *polnyj* 'full', *(polnym)-polno* 'a lot', lit. '(fully) full'.

(56)   Russian
*Primerov,   k nesčast'ju,   **polnym-polno**.*   (RNC)
example:GEN.PL to misfortune:DAT.SG fully-full
'Examples, unfortunately, are **numerous**.'

A similar semantic development occurs in lexemes within the domain of IN-TACTNESS. We find terms for 'whole' participating in a quantitative construction, similarly to lexemes for 'full' and carrying a similar component of evaluation: the amount is considered large.

(57)   a.   Spanish
*Cuánt-a   agua beb-ió?   – Un   vaso **enter-o**.*
how.much-F water drink-AOR.3SG – INDF.M.SG glass whole-M.SG
'How much water did he drink? – A **whole** glass.'
b.   Russian
*Ja poexala   na sklad   i   kupila   **celyj***
I   go:PST.F.SG on warehouse:ACC.SG and buy:PST.F.SG whole:ACC.M.SG
gruzovik   dosok.   (RNC)
truck:ACC.SG board:GEN.PL
'I went to a warehouse and bought a **whole** truckload (lit. 'a whole truck') of planks.'

A truck may have been used to deliver the planks in the situation described in (57b) but obviously was not sold along with the planks. Rather, the truck is mentioned only as a measure of quantity.

3.3.3   *Intensifiers*

There are usages where the terms of fullness no longer evoke an image of a container, and the only remaining component of meaning is the idea of a maximum.[7] Eventually these maximizers can evolve into a simple intensifier for abstract concepts. A common trait of these usages is that they preclude overt syntactic expression of contents.

Terms for COMPLETENESS also commonly develop the metaphorical meaning of the maximum or the greatest degree. Some languages use terms for 'full' and others use terms for 'complete' or 'whole/entire' in the same situational frames, ranging from Russian where almost all of the situations are covered by a lexeme for 'full', to Mandarin where a lexeme for 'complete' is preferred.

(58) a. Serbian
*Održavaj   kurs,   **puna**   snaga!*
maintain:IMP.SG course:ACC.SG full:NOM.F.SG power:NOM.SG
'Maintain course, **full** speed ahead!'   (Glosbe_SR)

b. Spanish
El   capitán de-l   barco estaba   dirigiendo   la
DEF.M.SG captain of-DEF.M.SG ship be:IPF.3SG steer:PTCP.ACT DEF.F
*nave a **plen-a**   velocidad.*
nave to full-F.SG speed
'The captain of the ship made it go at **full** (i.e. maximum) speed.'

c. Spanish
acceso completo 'full (lit. 'complete') access', *victoria completa* 'complete victory'

d. Russian
polnaja skorost' 'full speed', *polnaja otvetstvennost'* 'full responsibility', *polnaja pobeda* 'complete (lit. full) victory'

e. Mandarin
quán sù 'full (lit. 'complete') speed', *quán zé* 'full (lit. 'complete') responsibility', *quán shèng* 'complete victory'

---

7.   Two relatively less abstract, physical phenomena which are sometimes described as 'full' are the maximum phases of the moon and tide. Interestingly, in such unrelated languages as Mandarin and Spanish 'full moon' (*mǎn yuè, luna llena*) is an exception to the way a maximum is usually expressed. Mandarin normally uses *quán* 'whole', and Spanish prefers *pleno* 'full', while *lleno* is primarily used to describe physical containers. The same is true of 'full tide' (*mǎn cháo* in Mandarin; *marea llena* exists in Spanish, even though *marea alta* 'high tide' is more frequent).

The Mandarin root *quán* 'complete' and some of the compounds made with it are also borrowed into Japanese as *zen*, e.g. **zen-ryoku** 'all of one's strength'. Not all abstractions in Mandarin are covered by *quán*. The main physical term for 'full', *mǎn*, can be seen in such combinations as **mǎn fēn** 'full marks' and **mǎn yuán** 'full staff', and, as a loan-morph, in their respective Japanese counterparts **manten** and **man'in**.

In Japanese the loanword *furu* from English *full* is also used:

(59) Japanese
*Kōjō   ga   **furu** kadō   shite mo juyō   ni   oitsuk-anai.*
factory NOM full operation do   even demand DAT overtake-NEG
'Even running the factory **all-out**, we can't keep up with demand.'

(Kenkyusha)

As a development of the idea of the upper limit, we also find lexemes for 'full' modifying predicates to describe extent or degree:

(60) a. Mandarin
Liǎng tiān bàn de   shíjiān liúyán yǐjīng xiěmǎn le   3 běn.
two   day half ATR time   message already write.full MOD 3 CLF
'In two and a half days [the visitors'] impressions have already filled (lit. 'write-**filled**') 3 notebooks'.   (CCL)

b. Khanty
*χɔt   pitər **teλijewa/teλ** tăχər-man*
house wall full   hang-CVB
'a wall **fully** hung [with things]'

Russian expresses a similar maximum extent with the adverb *polnost'ju* 'completely', lit. 'in fullness', which is derived from *polnyj* 'full', and Serbian – with the adverb *potpuno* derived from *pun* 'full'.

(61) Serbian
***Potpuno** sam   nevin.*   (Glosbe_SR)
fully   be:PRS.1SG innocent:NOM.M.SG
'I am completely (lit. '**fully**') innocent'.

4. Conclusion: Semantic maps

In conclusion, we give some provisional semantic maps for the domain (for more on semantic maps in lexical typology, see Chapter 1; for a recent general review see Georgakopoulos & Polis 2018).

The core part of the subdomain of EMPTINESS is formed by five frames which reflect (1) the opposition of containers vs. non-container objects, (2) the shape of the container (deep or flat), and (3) the temporary or long-term absence of contents in the container; in our data permanent absence is only lexicalized when the contents are animate, i.e. inhabitants (see more on such correlations between parameters in Chapter 1). We propose the following frames:

- hollow object
- deep container
- flat container
- currently unoccupied container
- uninhabited location

Note that the set of frames is fully data-driven; that is, we introduce only frames that are lexicalized in at least one of the languages surveyed. It is possible that a larger sample may demand adding parameters or reformulating some of the existing ones.

We also do not treat the kind of contents (people vs. objects) as an independent parameter, since we have found that in our data the animacy of the object is conflated with the parameter of long-term absence. Yet our experience in typological analysis of other qualitative domains suggests that this kind of taxonomic opposition may be relevant for the lexicalization of the domain of EMPTINESS as a whole (see Chapter 1 for detail).[8]

The resulting frame system can be visualized as in Figure 1. On the maps we use capital letters for frames and lowercase italics for lexemes.

```
              EMPTY (SHALLOW) ----------------- UNOCCUPIED

                            |


HOLLOW  -------------      EMPTY (DEEP)


                            |


                      UNINHABITED
```
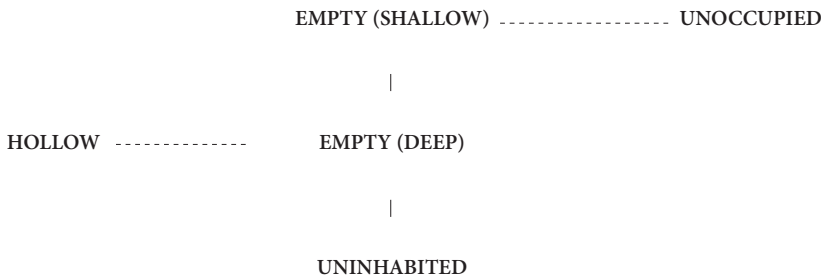
Figure 1. Semantic map of the EMPTY domain

---

8. In particular, the Serbian term *pust*, which at present is the preferred term for long-term absence of inhabitants, historically used to mean the absence of people irrespective of duration. This formed an opposition with *prazan*, the term for the absence of inanimate contents.

In our sample, English illustrates systems which distinguish between the frames 'empty of contents' and 'hollow' but do not distinguish temporary absence of contents versus long-term absence of inhabitants; the structure formed by the Chinese, Japanese, and Spanish terms in the domain of EMPTINESS is basically similar.
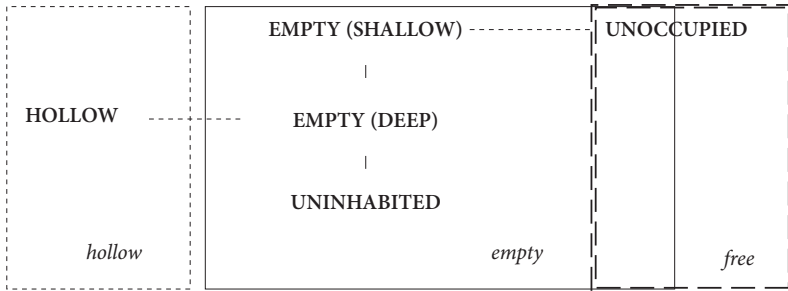
Figure 2. Semantic map of the EMPTY domain: English

Russian has a distinct term for 'hollow', *polyj*, and a broader term *pustoj* for 'empty of contents'. *Pustoj* 'empty of contents' tends to be used instead of *polyj* 'hollow' and possibly may replace it with time.
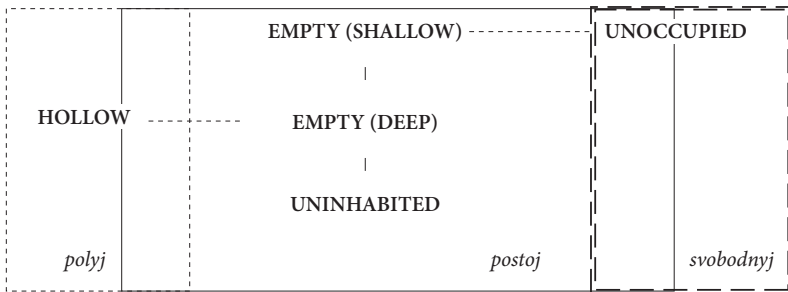
Figure 3. Semantic map of the EMPTY domain: Russian

Serbian makes a distinction between the absence of temporary occupants (e.g. "a free table") vs. long-term inhabitants (e.g. "an uninhabited island"); the former situation is described by the general adjective *prazan* 'empty', and the latter by the specific term *pust*.
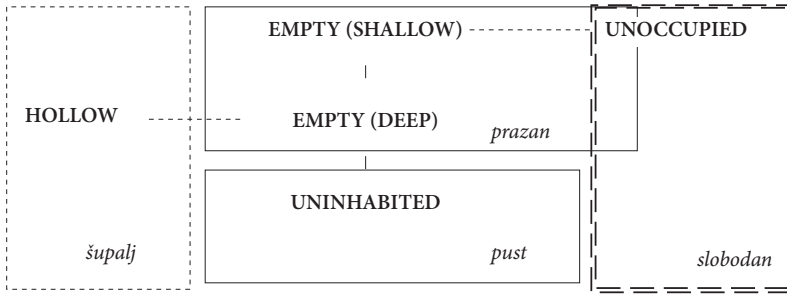
Figure 4. Semantic map of the EMPTY domain: Serbian

Korean has a unique opposition of shallow vs. deep containers and no specific term for 'hollow' – hollow objects are described by the adjective for 'empty and deep'. The main term for emptiness is the one for shallow containers, *pita*.
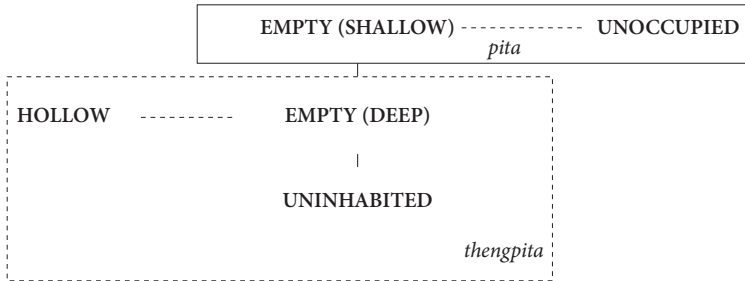
Figure 5. Semantic map of the EMPTY domain: Korean

Semantic maps not only allow us to compare the scope of meanings of lexical items across languages, but give a clear picture of the asymmetry of the antonymous domains FULL and EMPTY.

The features of 'deep/flat container' and 'permanent/current absence of inhabitants', relevant to the domain of EMPTINESS, have no parallels in the domain of FULLNESS. The domain of FULLNESS, in its turn, introduces the 'degree of fullness' (the opposition of ordinary, functional fullness vs. technical fullness to the limit) as a distinct opposition; the latter is often expressed by terms for overflowing liquids. Another factor relevant only to the FULL domain is the mereological opposition of substances or uncountable masses vs. limited sets of countable elements. Finally, there can also be specific means of expressing a full container as a measure of quantity. We believe that the frame system of the domain of FULLNESS includes the following frames, represented by the map in Figure 6:

– solid object
– functionally full container
– technically full container
– occupied container
– complete set
– quantitative use: full container as measure

```
            QUANTITATIVE          SET

                 |                 |

SOLID  -------------  FUNCTIONALLY FULL  ------------- OCCUPIED

                        |

                 TECHNICALLY FULL
```
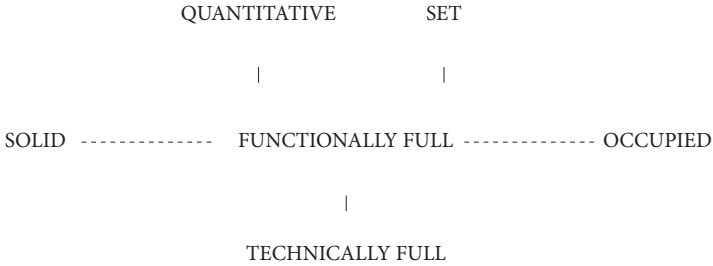
Figure 6. Semantic map of the FULL domain

English illustrates the opposition of functional vs. technical fullness, as well as competing terms for the quantitative construction and for sets.
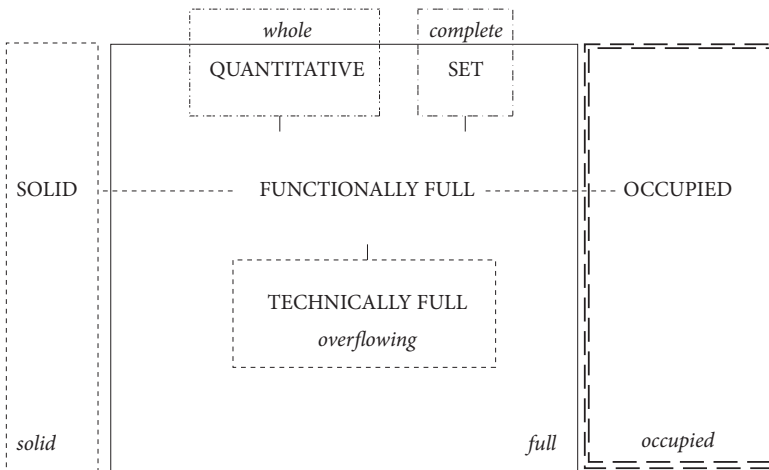
Figure 7. Semantic map of the FULL domain: English

In Serbian the main adjective for 'full containers' is widely used and even crosses over into the domain of AVAILABILITY.
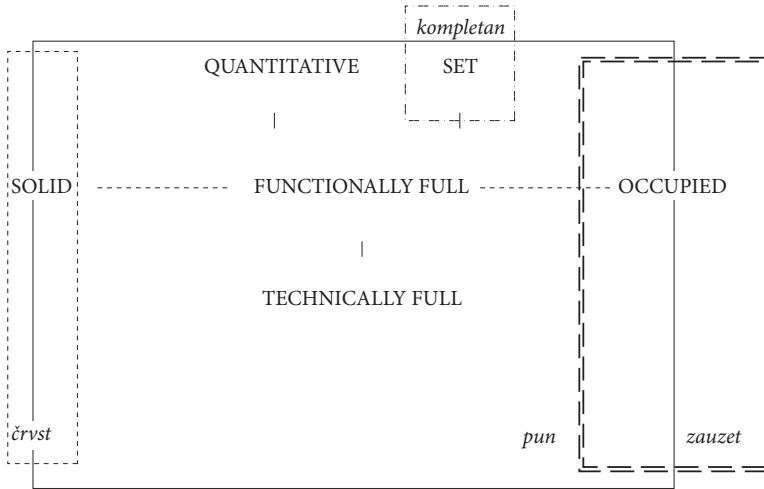
Figure 8. Semantic map of the FULL domain: Serbian

Spanish, besides having several terms for complete sets, uses a separate term in the quantitative construction.
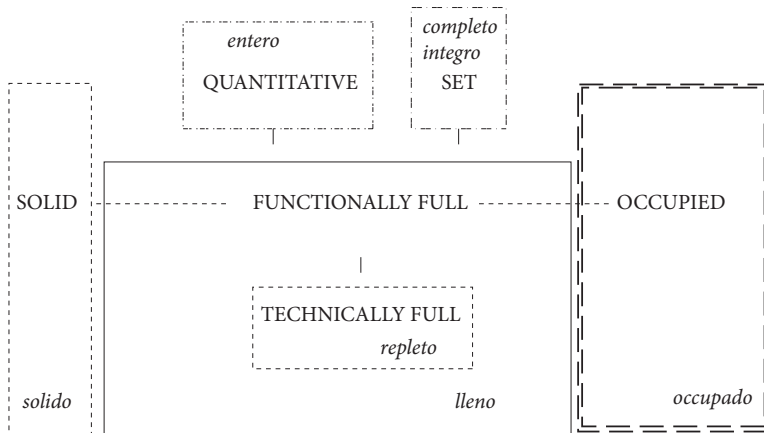
Figure 9. Semantic map of the FULL domain: Spanish

In Russian the adjective *polnyj* 'full' in the quantitative construction competes with a term from an adjacent domain, *celyj* 'intact', while *polnyj* 'full' covers some situations of completeness.
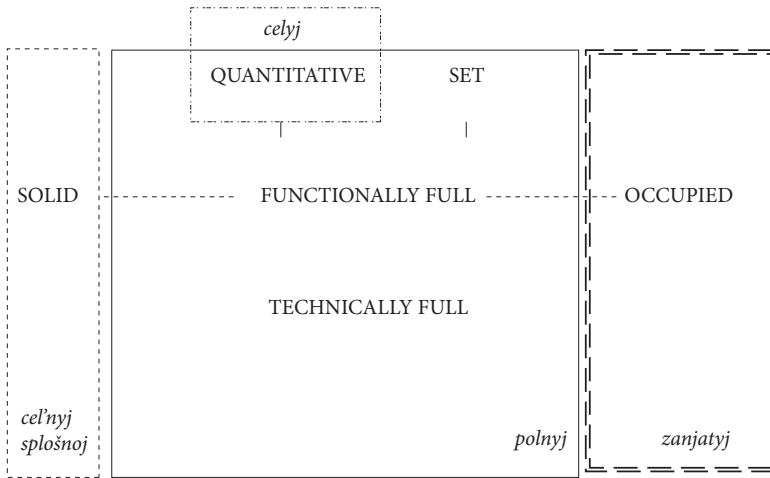
Figure 10. Semantic map of the FULL domain: Russian

The maps above show that, language-specific differences aside, the domains of FULLNESS and EMPTINESS in our sample have a roughly similar structure: there is a core term that describes the state of a container, and a few peripheral terms for some adjacent situations. The languages mostly differ in how far into these peripheral domains the core term may intrude.

Such structure of the domain is probably what determines the development of the metaphoric senses. In some other qualitative domains we can identify particular frames each of which give rise to a separate metaphoric shift (see, for example, the chapters on SHARP and SMOOTH in this volume). For FULL/EMPTY non-physical meanings are mostly found in the core terms, which cover several frames. Furthermore, similar semantic development can occur in adjacent domains, as when the idea of a maximum degree is expressed by terms of COMPLETENESS.

Usually metaphorization picks up the abstract concept behind a situation; the specific details are discarded and the relevant properties provide an image on which the new metaphoric meaning is based (cf. the concept of 'Generic Space' in the Conceptual Blending theory, Fauconnier & Turner 1998, 2003). These underlying images tend to differ for different frames or groups of frames within a domain. For example, as discussed in Chapter 6 of this volume, the domain of SMOOTHNESS has three separate frames which serve as starting points for metaphors, each with a different image: 'slippery', 'smooth' (both belonging to tactile perception), and 'flat' (visual perception). For 'slippery' the image is of a surface that does not provide reliable hold or footing – hence the shift to the meaning 'untrustworthy' (*a slick/ slippery person*). 'Smooth' focuses on the lack of surface irregularities as obstacles, conceptualized on the more abstract level as the absence of defects or difficulties

(*smooth speech*). For 'flat' the relevant image is uniformity, hence the metaphors of regular distribution (*evenly spaced desks, three even amounts*). Thus the different focus in the situations behind each term in the domain of SMOOTHNESS motivates how their metaphoric meanings develop in different directions. Similar mechanisms of metaphorization have also been observed outside qualitative domains, e.g. in terms of motion in liquids (see Rakhilina 2007).

The domains of FULLNESS and EMPTINESS appear to deviate from this picture, because semantic extensions cannot be clearly traced to individual frames, and even the terms from adjacent domains produce highly similar secondary meanings. In our opinion, this is due to the fact that all the frames within FULL and EMPTY have a similar underlying image. In the domain of EMPTINESS it is the lack of something that is normally present in this place; this is the core image behind both the terms for empty containers and for cavities in a solid object. In the domain of FULLNESS the image common to all frames is the upper limit, be it the maximum capacity of a container, or the completion of a set. From this point of view, the similarity of the metaphoric shifts is understandable – the shifts highlight the underlying similarity in the abstract models of the respective situations.

Of course, a thorough typology of semantic shifts cannot be created without a sample much larger than the one used in our current attempt (cf. Chapter 2). It is likely that new data will show more patterns in which secondary meanings of FULL and EMPTY are formed and organized.

Abbreviations

1, 2, 3	the 1st, 2nd, 3rd person	DEF	definite
ABL	ablative	F	feminine
ACC	accusative	GEN	genitive
ADD	additive particle	IMP	imperative
ADV	adverbializer	INDF	indefinite
AOR	aorist	INF	infinitive
ATR	attributivizer	INS	instrumental
BEN	benefactive	IPF	imperfect
CLF	classifier	IRR	irrealis
CMT	comitative	LOC	locative
COMP	comparative	M	masculine
CONJ	conjunction	MOD	modal particle
CONT	continuous	N	neutral
CVB	converb	NEG	negation
DAT	dative	NOM	nominative

NPST	non-past	QUOT	quotative
PASS	passive	REFL	reflexive
PL	plural	REL	relative pronoun
PRS	present	SG	singular
PST	past	TOP	topic
PTCP	participle		

References

Bhat, D. N. S. 1999. Intransitive / adjectival predication (review article). *Linguistic Typology* 3(1): 111–125.

Fauconnier, G. & Turner, M. 1998. Conceptual integration networks. *Cognitive Science* 22(2): 133–187. https://doi.org/10.1207/s15516709cog2202_1

Fauconnier, G. & Turner, M. 2003. *The Way We Think*. New York: Basic Books.

Georgakopoulos, Th. & Polis, S. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2): 1–33. https://doi.org/10.1111/lnc3.12270

Goddard, C. & Wierzbicka, A. 2007. NSM analyses of the semantics of physical qualities: *sweet, hot, hard, heavy, rough, sharp* in cross-linguistic perspective. *Studies in Language* 31(4): 765–800. https://doi.org/10.1075/sl.31.4.03god

Grady, J. 2007. Metaphor. In D. Geeraerts, H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*: 187–213. Oxford: Oxford University Press.

Heine, B. 1997. *Possession: Cognitive Sources, Forces and Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511581908

Herskovits, A. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge: Cambridge University Press.

Kustova, G. I. 2006. Valentnosti i konstrukcii prilagatel'nyx [Arguments and constructions of adjectives]. In *Komp'juternaja lingvistika i intellektual'nye tehnologii. Trudy meždunarodnoj konferencii "Dialog-2006"* [Computational Linguistics and Intellectual Technologies. Papers from the International Conference "Dialog-2006"], 323–328. Moscow: RGGU.

Lakoff, G. & Johnson, M. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Rakhilina, E. V. 2007. Tipy metaforičeskix upotreblenij glagolov plavanija [Types of metaphorical uses of aqua-motion verbs]. In *Glagoly dviženija v vode: leksičeskaja tipologija* [Verbs of AQUA-Motion: Lexical Typology], T. A. Maisak & E. V. Rakhilina (eds), 76–105. Moscow: Indrik.

Tagabileva, M. & Kholkina, L. 2010. Kačestvennye priznaki 'pustoj' i 'polnyj' v tipologičeskom osveščenii [Qualities 'full' and 'empty' in typological perspective]. *Acta Linguistica Petropolitana* 4(3): 167–169.

Tagabileva, M., Kholkina, L. & Kiryanov, D. 2013. Semantic domains 'full' and 'empty': a cross-linguistic study. In *Association for Linguistic Typology. The 10th Biennal Conference (Leipzig, 2013 August 15–18). Abstracts*, 178–179. <http://www.eva.mpg.de/lingua/conference/2013_ALT10/pdf/abstracts/Abstracts_ALT10_complete.pdf>

Taylor, J. 1989. *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Tolstaya, S. M. 2008. *Prostranstvo slova. Leksičeskaja semantika v obščeslavjanskoj perspektive* [Extent of a Word. Lexical Semantics in a General Slavic Perspective]. Moscow: Indrik.

Vandeloise, C. 1991. *Spatial Prepositions: A Case Study from French*. Chicago: Univ. of Chicago.

Vandeloise, C. 1994. Methodology and analyses of the preposition. *Cognitive Linguistics* 5(2): 157–184. https://doi.org/10.1515/cogl.1994.5.2.157

Sources

BCCWJ – *The Balanced Corpus of Contemporary Written Japanese*, <www.kotonoha.gr.jp/shonagon/>

BCS – *Bosnian, Croatian, Serbian. A Textbook*. Wisconsin: The University of Wisconsin Press, 2010.

CCL – *Peking University CCL Online Corpus*, <http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai>

CE – *Corpus del español*, <http://www.corpusdelespanol.org>

Collins – *Collins Spanish-English Dictionary*, <https://www.collinsdictionary.com/dictionary/spanish-english/>

CREA - *Corpus de Referencia del Español Actual*, <http://corpus.rae.es/creanet.html>

Diccionario Salamanca de la lengua española [The Salamanca Dictionary of Spanish]. Barcelona: Santillana, 1996.

EsTenTen18 – *Spanish Web Corpus on the Sketch Engine Platform*, <https://www.sketchengine.eu/estenten-spanish-corpus/>

Glosbe_ES – *the multilingual online dictionary and translation memory. Spanish – English*, <https://glosbe.com/es/en/>

Glosbe_SR – *the multilingual online dictionary and translation memory. Serbian – English*, <https://glosbe.com/sr/en/>

Kenkyusha – Kenkyusha's New Japanese-English Dictionary. 5th edition. Tokyo: Kenkyūsha, 2003.

Korean-English Dictionary. Seoul: YBM/Si-sa, 2006.

Narumov B. P. *Ispansko-russkij slovar'* [Spanish-Russian Dictionary]. Moscow: Russkij jazyk, 1988.

Naver Korean-English Dictionary, <https://korean.dict.naver.com/english/main.nhn?sLn=kr>

Park Hyeng Gyu. *Korean-Russian dictionary*. Seoul: Korean-Russian Cultural center, 2006.

RNC – *Russian National Corpus*, <https://ruscorpora.ru/new/en/search-main.html>

Sejong – *The Sejong Corpus*, <http://www.sejong.or.kr>

Turover G. Ja. *Bol'šoj russko-ispanskij slovar'* [Unabridged Russian-Spanish Dictionary]. Moscow: Russkij jazyk, 2003

Xiàndài Hànyǔ cídiǎn [Dictionary of Modern Chinese]. 6th edition. Běijīng: Shāngwù yìnshū guǎn, 2012

Typology of dimensions

Alexey Kozlov[1,2] and Mariia Privizentseva[3]
[1]HSE University, Moscow / [2]Institute of Linguistics, Russian Academy
of Sciences / [3]Leipzig University

The paper focuses on the lexical typology of dimensional terms such as English *long*, *deep*, *wide*, etc. Compared to other semantic fields, this one is relatively well-studied; however, the present study is the first to approach it from the modern typological point of view. We propose a semantic map of dimensional terms, which outlines the possible and impossible colexification patterns in the domain. However, other regularities appear likely to exist, which cannot be captured by the model of semantic mapping. We discuss the potential restrictions on colexifications, and suggest explanations for them.

Keywords: dimensional terms, lexical typology, semantic maps, spatial expressions

1. Introduction

> Arma virumque cano, Troiae qui primus ab oris
> Italiam, fato profugus, Laviniaque venit
> litora, multum ille et terris iactatus et **alto**
> vi superum saevae memorem Iunonis ob iram;
> multa quoque et bello passus, dum conderet urbem,
> inferretque deos Latio, genus unde Latinum,
> Albanique patres, atque **altae** moenia Romae.

In the beginning of Vergil's Aeneid, the reader encounters two occurrences of the lexeme which we may represent as *altus.* The first of them is a substantivized form, which occurs in conjunction with the noun *terris* 'land-ABL.PL' (or rather, with the noun *mare* 'sea.ABL', a frequent collocate of *altus*, which is omitted in this verse):

(1) *multum ille et terr-is iacta-t-us et alt-o*
    much he and land-ABL.PL throw-PTCP.RES-NOM.SG and *altus*-ABL.SG'
    'He has been for a long time thrown (by the storm) between the land and the deep [sea/places].'

Aeneas was thrown by a storm hither and thither throughout the sea, from the land to the deep places – and 'deep', or 'a deep place', is exactly what *altus* means there. This meaning is fairly frequent for this word: along with *altum mare* we find *altus puteus* 'deep well', *altissima flumina* 'the deepest river', *altum vulnus* 'deep wound', etc.

However, in the end of the passage, within the noun phrase *alta moenia Romae* 'the high walls of Rome', the word *altus* yields quite a different interpretation – or at least a different English translation. Such uses of this word are even more common than those of the previous kind, as in the examples *altae parietes* 'high walls, *altus acervus* 'high pile', and *alta turris* 'high tower'.

Such polyfunctionality might strike speakers of modern European languages as unexpected and bizarre, as their languages consistently lexicalize these two meanings as separate words. However, combining them in a single lexical item is not an idiosyncratic property of Latin; the same pattern is attested, for example, in Izhma Komi (Permic, Uralic), where the two concepts, 'high' and 'deep', are expressed by the same word *džudžɨd*.

The present article deals with cross-linguistic variation in the meaning of terms such as *high*, *deep*, *thick*, or *wide* and the like. They describe the size of a physical object in one of its dimensions, so we will further refer to them using the umbrella expression *dimensional term*.

The methodology we exploit in this study draws on the set of methods utilized by the MLexT group (see Chapter 1 in this volume). Our primary dataset was obtained by eliciting answers from native speakers who were asked to fill in a questionnaire; in the majority of cases, several speakers were interviewed. The original questionnaire was designed according to our provisional hypotheses about what could be relevant for the semantic field under examination. However, sometimes in the process of data collection we would encounter new, previously unattested semantic distinctions; each time we came across such a distinction, we modified the questionnaire accordingly and collected the lacking data from the languages already represented in the sample. After our speakers translated the questionnaire, for each sentence we asked whether the speaker could felicitously replace the chosen dimensional term with some other term, either a term used by the speaker in other sentences or a term supplied by dictionaries of the relevant language. Whenever necessary, we requested a consultation from experts on these languages.

The entries in our questionnaire are English (or, when working with the languages of Russia, Russian) sentences containing noun phrases with dimensional terms. The final version of the questionnaire contains several examples illustrating each frame of the semantic map of dimensions (see Section 2).[1]

---

1.   The full version of the questionnaire is available online at lextyp.org.

To verify the speakers' judgements, we used English, German, and Russian corpora. The other languages of our sample, including the minor languages of Russia are under-resourced; corpora are either unavailable or too small and unrepresentative to draw any specific conclusions about the semantics of their dimensional terms. We also consulted dictionaries of the respective languages, but dictionary data is rarely sufficient or precise enough for our purposes. Therefore, we mostly relied on the judgements of the speakers.

Considering this limitation, our sample is a choice of convenience; it includes only languages for which we could obtain sufficient reliable data. We examined the following languages and families:

- Indo-European: German, East Armenian, English, Irish, Italian, Lithuanian, Latvian, Russian, Serbian, and Spanish;
- Uralic: Beserman Udmurt, Finnish, Izhma Komi, Hungarian, Moksha, Tundra Nenets, Northern Khanty, Mari (Hill and Meadow);
- Altaic: Azeri, Buryat, Kazakh, Kirghiz, Tatar, Turkish;
- Northeast Caucasian: Aghul, Avar, Itsari Dargwa, Ingush, Lezgian, Tabasaran, Udi;
- Akebu (Kwa, Niger-Kongo), Georgian (Kartvelian), Chukchi (Chukotko-Kamchatkan), Ch'orti' (Mayan), Kabardian (Northwest Caucasian), Modern Hebrew, and Mandarin Chinese.

As can be seen, our sample is heavily biased towards Northern Eurasia. For a preliminary study like ours, this does not appear to pose a grave problem, as we intend to describe at least some typologically relevant features and not to offer an exhaustive universal inventory.

The structure of the article is as follows. In Section 2.1 we introduce the central notion of the subsequent discussion, that of topological class. Section 2.2 proceeds to establish the list of atomic meanings, or frames, which is an indispensable component of cross-linguistic lexical semantic comparison. In Section 2.3 we come up with the semantic map which organizes the frames that emerge from the data and generates hypotheses about general cross-linguistic restrictions on the meanings of dimensional terms. Sections 3, 4, and 5 offer a detailed description of the attested patterns of colexification in the languages of the sample. Section 6 sums up the data presented in the previous sections and addresses the two approaches to the semantics and typology of dimensional terms that gained prominence in the previous decades: that by M. Bierwisch and E. Lang (Bierwisch & Lang 1989; Lang 2001), and by Anna Wierzbicka (Wierzbicka 2006); we discuss them in the light of the typological data presented in this article.

2.   The semantic map of dimension

2.1   Topological classification

A particular application of frame semantics is adopted in this study, in line with the other MLexT projects (the detailed discussion of how the notion of frame introduced by Fillmore 1978, 1982 can be used in lexical typology, see Chapter 1 in this volume). We do not attempt to analyse the meanings of the dimensional lexemes into components; rather, we seek to identify a set of frames, i.e. the prototypical contexts or situations in which the lexeme can be used. Then we plot these frames on a semantic map, exactly like grammatical typologists do when they arrange grammatical functions on a map (see, e.g., Haspelmath 2003). The meaning of a lexeme is thus regarded as a cluster of its "atomic meanings", i.e. of its frames on a map.

The prototypical situations in the questionnaire, each of which is strictly defined, serve as the *tertium comparationis* in our typological study. For each language of our sample, we establish the terms, or lexemes, that lexicalize each of the "atomic meanings". Before we present the list of these meanings, let us first introduce the notion of *topological class*.

Consider an arbitrary physical object that can be rotated around its own axis, turned upside down, and viewed by a human observer from an arbitrary point and angle of view. If you wanted to describe it by means of a dimensional term, how could you decide which of its dimensions to pick to characterise the object?

Obviously, the dimension that you are going to speak about should be salient or distinguishable regardless of the position of both the object and the observer. One should be able to differentiate it on the basis of the inherent spatial properties of the object, i.e. its shape and size.

If we say (in English) that an object is *thick,* we are thereby making certain assumptions about its form, namely, that this object must have a dimension it can be thick in. By saying so, the speaker implies that the object is a member of a certain class of objects: either layer-like flat things like *paper*, *fabric*, or *ice*, or longish objects, e.g. *a branch of a tree*, *a pencil, or a string*. Such an assessment about the form of an object is the kind of information that is necessarily carried by every dimensional term.

Let us note that in the languages of our sample, dimensional terms typically combine with a relatively small subclass of object nouns. Indeed, no dimensional term can modify a noun like *sphere*, as all the dimensions of a sphere are equal to each other, or words like *goose* or *headphones*, which denote objects that are highly complex and elaborate in shape. In order to license a dimensional term to describe it, an object should belong to a certain *topological class*.

The notion of topological class is based on the idea that language provides a naïve geometrical classification of (some) physical objects (cf. Talmy 1983; Jackendoff & Landau 1993). In all the languages of our sample, the objects described with dimensional terms demonstrate a distinct tendency to agglomerate around a small number of prototypical shapes.

Language categorization is never entirely categorical; the subtlety of subdivisions can increase with the depth of analysis. According to a widespread view, semantic (or participant) roles is another possible example of such elusiveness; Dowty (1991) claims that the semantic roles of each verb's arguments are unique, and the degree of granularity depends on the researcher's objectives. The situation with topological classes is presumably alike, as the variety of forms and shapes of individual objects is virtually limitless. Nevertheless, some distinctions turn out to be typologically relevant, while others are not.

The topological classification we rely on in this study emerges from our data. Let us look at a toy example of how it works. In Kazakh, there are two adjectives, *ensiz* and *tar,* both of which are translated into English as *narrow.* However, they are rarely interchangeable: except for a minor subset of cases (on which we will elaborate later), they occur in a complementary distribution with each other. *Tar* is mostly used to describe narrow parts of space: passages, tubes, doors, windows, paths, and roads. *Ensiz* primarily applies to flat and long artifacts: ribbons, stripes on a dress, wooden boards etc., as well as to some natural objects which resemble them, as *a narrow stripe of land* (especially on a map).

(2) KAZAKH
    a. *ensiz (*tar) taqta*
       narrow   board
    b. *tar (*ensiz) žol*
       narrow   road
    c. *tar (*ensiz) tesik*
       narrow   hole

Tundra Nenets (Samoyedic, Uralic) manifests another distinction in this zone. Both of the two adjectives, *tiya* and *pik*, are also translated into English as *narrow*, and are (almost) complementarily distributed. However, they divide the semantic space in a different way. *Pik* denotes objects with a hollow space inside, and *tiya* combines with all types of flat and elongated objects, no matter whether they are artificial or not. Thus, 'narrow (board)' and 'narrow (hole)' are described differently in Nenets similarly to Kazakh, but unlike in Kazakh, roads tend to pattern with boards rather than with holes: this is shown in (3).

(3)   NENETS

    a.   *tiya (*pik) latə*
        narrow   board

    b.   *tiya (*pik) sexeri*
        narrow   road

    c.   *pik (*tiya) s'i*
        narrow   hole

The Nenets and the Kazakh data take us to the following conclusion: from a cross-linguistic point of view, there are at least three classes of objects which are relevant for terms like *narrow* – firstly, flat and elongated inanimate objects; secondly, roads and pathways; and thirdly, hollow spaces. This is how we arrive at the list of classes, which is the first result of our study. For the sake of convenience we will refer to our classes by labels, e.g. the three aforementioned groups will be labelled as *Stripes*, *Roads*, and *Tubes & Holes*. We chose the labels that reflect certain characteristic properties of the corresponding classes, but certainly we do not commit ourselves to all the implications that may follow from the choice of particular English words used as labels.

2.2   The list of classes

We now present the list of classes which, according to our data, are relevant for lexicalization of dimensions in languages of the world. We have just seen the toy proofs of the distinction between Stripes vs. Roads vs. Tubes & Holes on the basis of the Nenets and the Kazakh data; the other distinctions emerge from our data in a similar way (the data will be presented below). For each class, we propose a semantic definition, present several examples, and cite the English dimensional terms that can describe the words denoting objects of these classes.

–   **Pivots** are elongated objects in which one of their dimensions is considerably larger than the second and the third ones (Figure 1). For an observer and / or a user, the difference between the two latter dimensions is less important. Prototypical pivots are branches of trees, human fingers, ropes, etc.: they are round or nearly round in the cross-section. Pivots can be either *long* or *short*, and either *thick* or *thin*.
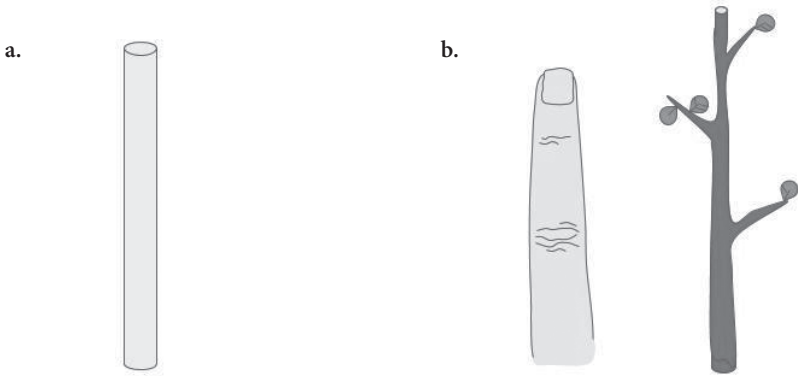
Figure 1. Types of Pivots

–   **Poles** (Figure 2) are very similar to Pivots except one crucial difference: Poles have a fixed upright orientation. The most illustrative examples of Poles are trees, guideposts, columns, and humans, which can be *tall* or *short*, and *thick* or *thin*. (In the remainder of the paper, we will use the term Pivots to refer only to non-vertically oriented Pivots, i.e., such Pivots which are not Poles.)
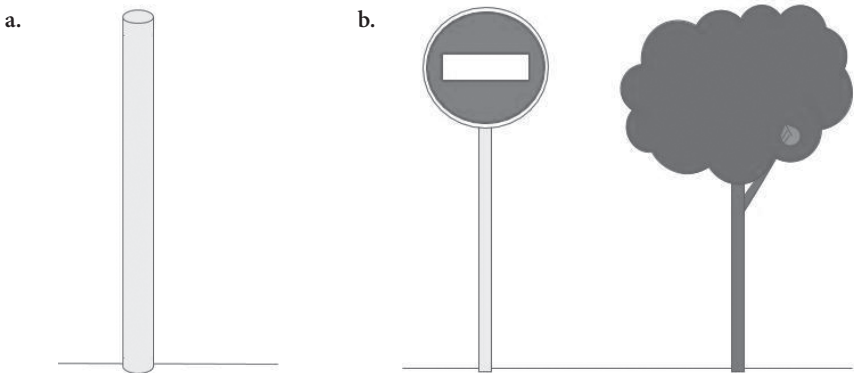
Figure 2. Types of Poles

–   **Barriers** are vertically oriented objects that are flat (Figure 3). This class embraces different kinds of walls, fences, and ramparts, as well as buildings which are longer horizontally than vertically, and various other objects which have a vertical orientation but do not have a pronounced pole-like shape. Barriers can be *high* or *low*, and *thick* or *thin* (and sometimes also *long* or *short*).

a.

b.
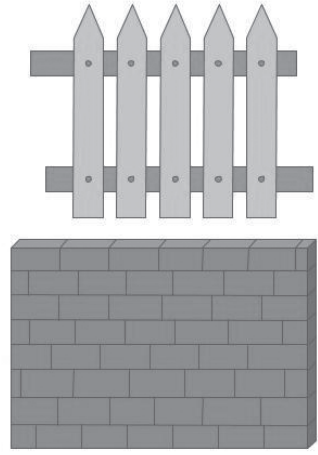
Figure 3. Types of Barriers

– **Layers** (Figure 4) are flat objects either with a well-defined border (book, board, or blanket) or without one (paper, ice, cloth). Generally, they can only be *thick* or *thin*.

a.
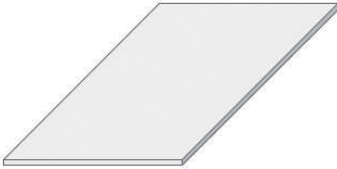
b.
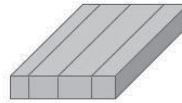
Figure 4. Types of Layers

– **Penetrable Layers** (Figure 5) is a small yet important subclass of Layers. Here belong the Layers that consist of some sticky substance one can plunge into, e.g. snow, mud, or sand. They can be *deep* or *shallow*.

a.

b.
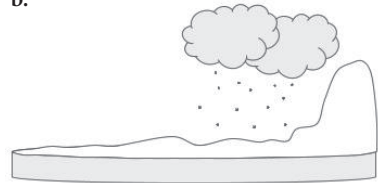
Figure 5. Types of Penetrable Layers

– **Stripes** (Figure 6) differ from Pivots in only one respect: stripes are flat, i.e. of their three dimensions, the smallest one either does not have a salient enough dimension or is considerably smaller than the second one. Examples include ribbons, stripes on a flag or on a dress, wooden boards, etc. They are not necessarily artifacts, e.g. *a stripe of land between the sea and the rocks* is also a Stripe. Stripes can be *long* or *short*, *thick* or *thin*, and *wide* or *narrow*.

a.　　　　　　　　　　　　b.
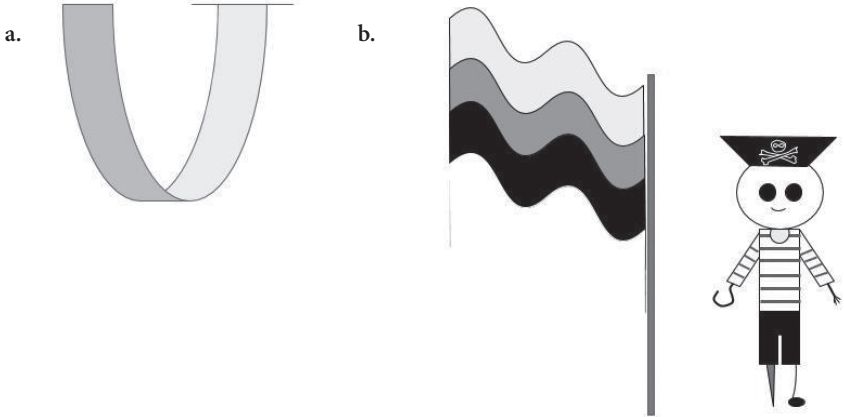
Figure 6. Types of Stripes

– **Roads** (Figure 7) are elongated spaces one can walk along: streets, rivers, paths, bridges, etc. They can be *long* or *short*, *wide* or *narrow*, but typically they are not *thick* or *thin*.

a.

b.
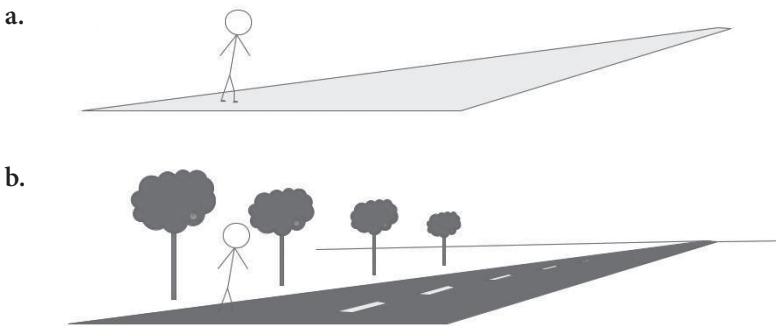
Figure 7. Types of Roads

– **Surfaces** (Figure 8) are places which are not, or do not necessarily need to be, elongated: fields, glades, yards etc. They can be *wide* (but cannot be *narrow*).

a.

b.
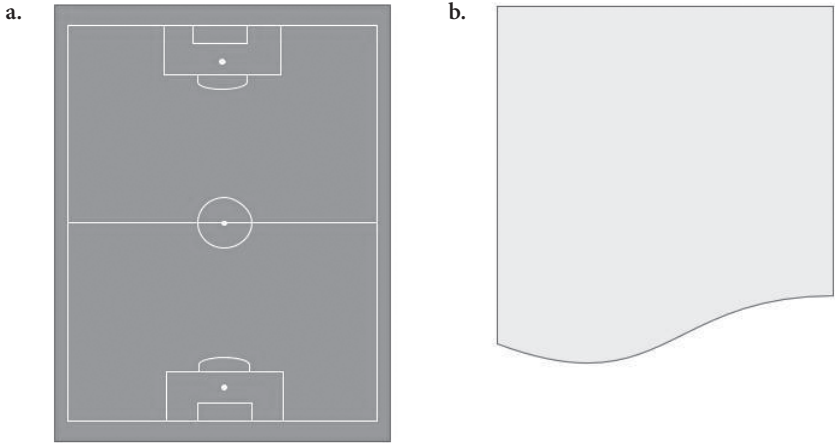
Figure 8. Types of Surfaces

– **Holes** (Figure 9) are empty spaces in layer-like objects: holes in walls, in a floor, etc. Important members of this class are doors and windows. Holes (unless they are square-shaped) have only one dimension, in which they measure as *wide* or *narrow*.

a.

b.
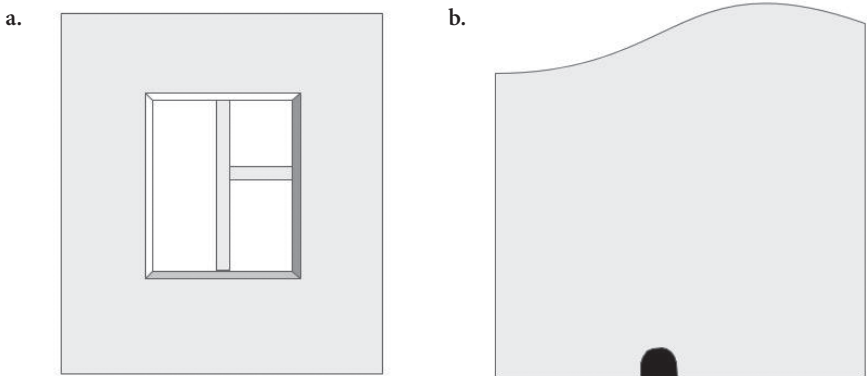
Figure 9. Types of Holes

– **Tubes** (Figure 10) are elongated empty spaces, like pipes, chimneys, corridors, and burrows. They can be *wide* or *narrow*, and *long* or *short*. In the case of Tubes & Holes, the dimension in which they can be wide or narrow is defined by their inner (rather than the outer) boundaries.

a.

b.
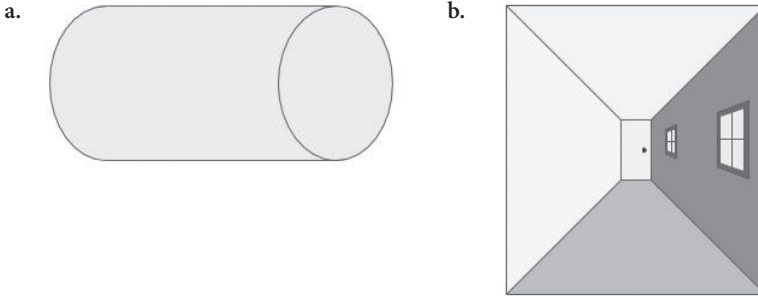
Figure 10. Types of Tubes

– **Pits** (Figure 11) are vertically oriented holes in the ground or in other surfaces humans can stand on, e.g. ditches, trenches, coal mines, etc. They can be *deep* or *shallow* (and perhaps *long* or *short*, and *wide* or *narrow*, depending on their form).

a.

b.
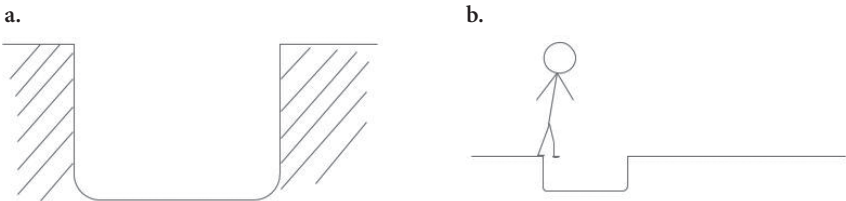
Figure 11. Types of Pits

– **Waterbodies** (Figure 12) are Pits filled with water: lakes, rivers, pounds, etc. In English, they can be described by the same adjectives as Pits.

a.

b.
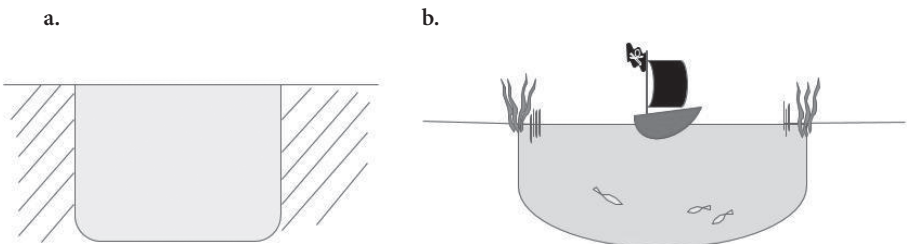
Figure 12. Types of Waterbodies

The reason why these topological classes prove to be cross-linguistically relevant for the meaning of dimensional terms is quite obvious: the members of each class feature some (and in the case of Stripes, all) dimensions that are discernible enough on a certain basis, either because of their inherent spatial characteristics, or due to the conventional scenarios of how humans interact with them. Later on in the detailed discussion of the data, we are going to present the evidence that justifies each of the classes listed above. Such objects as headphones or spheres lie outside of this classification simply because they have no dimension that would be considerably more salient than the two others.

Importantly, our classification does not emerge from just a handful of necessary and sufficient conditions, but rather allows of a certain degree of flexibility, which is natural for a prototype semantic structure. Various non-dimensional features that influence the attribution of an object to a class are numerous, and the borders between the classes are fuzzy. It is more appropriate to think of the suggested classes as focal points which determine the partition of the semantic field, rather than of real discrete sets with clear boundaries.

The categorization of objects as members of a certain topological class is not carried out according to their physical properties alone. Besides the shape of an object, its class largely depends on how people interact with it. A simple example is *river*: it is perceived as a Road when somebody sails or swims along or across it, and as a Waterbody (similarly to *lake, well,* etc.) when something sinks or moves under the water. However, this duality does not refute our classification, as all the principles imposed by the system are observed in both of these classes.

Furthermore, borderline cases are plentiful, when objects can be subsumed into two topological classes which are adjacent to each other on the semantic map. For example, our conceptualization of *layers of snow* depends on how probable it is that a person can sink into it. In many languages that have two different words for depth of Penetrable and Impenetrable Layers (e.g. *thick* and *deep* in English), snow by the porch would be *deep* rather than *thick*, and snow on a window sill or on a roof would be *thick*. In Kazakh, which, as we have already seen, distinguishes narrow Stripes vs. narrow Roads, and narrow Tubes & Holes *(ensiz* and *tar),* there is a small class of objects which can be viewed both as "pieces of a stripe" and as "places" inside which a human can be placed; thus they behave like Roads and Stripes at the same time – for example, bridges or different kinds of seats and benches.

2.3   Building a semantic map

We now arrive at defining the meanings of our dimensional terms. It is quite obvious that the meaning of a standard dimensional term consists of two components. Firstly, each term presupposes that the object that is being described belongs to a

certain topological class, and secondly, it asserts that one of the object's dimensions is either large or small. Consequently, we introduce the following labels (in small caps) for the atomic functions, which are equivalent to "frames" in the MLexT terminology. (We define "frame" as a class of situations that are covered by one lexeme in all the languages of a sample; for the discussion of how this definition is connected to the traditional definition of frame in frame semantics, see Chapter 1 in this volume) Note that the atomic functions (or "frames") combine a dimension with a class: WIDE STRIPE, NARROW STRIPE, WIDE ROAD, NARROW ROAD, WIDE HOLE, NARROW HOLE, etc.[2]

These atomic functions are comparative concepts (in the sense of Haspelmath 2010) and are not claimed to be projected automatically or necessarily onto the semantic representation of words in specific languages. The fact that the English word *thick* combines two frames on a semantic map does not imply that we consider *thick* to be polysemous. (In fact, this question is worth a special discussion, but for the reasons of space we are not going to address it here.)

The result of our research is visualized as the semantic map of dimensions, which we compiled with the aid of the procedures described in Chapters 1 and 2 in this volume. The map represents a graph whose nodes correspond to frames. The nodes are connected by "edges"; if a lexeme covers two frames, there should exist a path along the edges from one frame to the other such that each intermediary frame along the path also belongs to the denotation of the lexeme. To phrase it differently, each lexeme of our sample covers a continuous area on the map.

Our map consists of two parts which are almost symmetrical: one for the large sizes ('high', 'deep', 'wide' etc.), and the other for the small sizes ('low', 'shallow', 'thin' etc.). The dashed line divides our map into the two areas which will be subsequently described in Sections 3 and 5. Colexification of frames across this line is very rare; cases of such colexification will be described in Section 4. Hereafter we will dub the area above the dashed line (from PIVOTS up to P(ENETRABLE) LAYERS) ALTUS (from the Latin word meaning 'high'), and the area below the line (from I(MPENETRABLE) LAYERS to SURFACES) LATUS (from the Latin word meaning 'wide').

---

2.   Hereafter we will adhere to the following spelling conventions. Names of frames (which are our comparative concepts) will be in small caps (SHORT PIVOTS), and names of topological classes will be capitalized (e.g. Pivots). The lexemes of the languages of our sample will be in italic (*wide*). Also, we will sometimes use single quotes ('short') to indicate the meanings of the corresponding English words. Note that the latter are not our analytical tools in any respect, but just shorthands (e. g. "zone of 'wide'" replaces a longer "WIDE STRIPES, WIDE ROADS, and WIDE TUBES&HOLES"). Another shorthand that we use is, for example, to refer to terms that lexicalize one or several frames whose labels being with WIDE- as "WIDE-terms".

LONG PIVOTS   SHORT PIVOTS

TALL POLES   SHORT POLES

HIGH BARRIERS   LOW BARRIERS

DEEP PITS   SHALLOW PITS

DEEP WATERBODIES   SHALLOW WATERBODIES

DEEP P-LAYERS   SHALLOW P-LAYERS

— —

THICK I-LAYERS   THIN I-LAYERS

THICK PIVOTS   THIN PIVOTS

WIDE STRIPES   NARROW STRIPES

WIDE ROADS   NARROW ROADS

WIDE TUBES & HOLES   NARROW TUBES & HOLES

BROAD SURFACES

Figure 13. Semantic map of dimensions

The only exception from this symmetry is the frame BROAD SURFACES which has no counterpart in the subdomain of small sizes. Unlike the WIDE-terms, the NARROW-terms never apply to Surfaces, and this seems to be one of the major antonymical asymmetries in the domain of dimensions. Fields and lakes either can or cannot be described as *wide*, but they are never *narrow* (unless they are elongated, thus falling into the category of Stripes).

Let us consider two examples. Figure 14 shows how English dimensional lexemes *long, short, tall, low, high, deep, shallow, thick, wide, thin, narrow,* and *broad* cover the frames on the map.
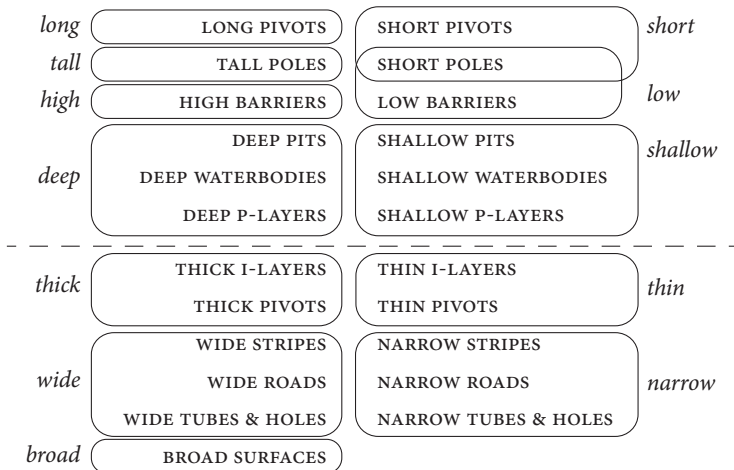
Figure 14. Semantic map of dimension lexemes: English

Note that the frame of SHORT POLES can be covered by both English *short* and *low*, which will be discussed in greater detail below. On the contrary, we take the frame SHALLOW P(ENETRABLE)-LAYERS to be lexicalized only by *shallow*, but not *thin* (as one may suggest): whereas collocations such as *shallow snow* or *shallow sand* are possible (e. g. in contexts such as *She tried to stick a pole into the shallow snow*), though infrequent, *thin snow seems to be infelicitous. *Thin layer of snow* is possible, but only by virtue of the fact that the word *layer* patterns with another topological class – Impenetrable Layers – rather than with Penetrable ones.

Figure 15 shows how Russian adjectives fit into the abstract semantic map.

dlinnyj	LONG PIVOTS	SHORT PIVOTS	*korotkij*
vysokij	TALL POLES	SHORT POLES	*nevysokij*
	HIGH BARRIERS	LOW BARRIERS	*nizkij*
glubokij	DEEP PITS	SHALLOW PITS	*neglubokij*
	DEEP WATERBODIES	SHALLOW WATERBODIES	*melkij*
	DEEP P-LAYERS	SHALLOW P-LAYERS	
tolstyj	THICK I-LAYERS	THIN I-LAYERS	*tonkij*
	THICK PIVOTS	THIN PIVOTS	
širokij	WIDE STRIPES	NARROW STRIPES	*uzkij*
	WIDE ROADS	NARROW ROADS	
	WIDE TUBES & HOLES	NARROW TUBES & HOLES	
	BROAD SURFACES		

Figure 15. Semantic map of dimension lexemes: Russian

In our analysis of the Russian data we mostly follow the approach of Rakhilina (2000), with the exception of minor details. For example, as the map in Figure 15 demonstrates, we consider *nizkij* not to be a full antonym of *vysokij* because collocations such as *nizkaja berëza* 'short birch' are ill-formed; only *nevysokaja bereza* is possible. Similarly, *melkij* is an antonym of *neglubokij* only for SHALLOW WATERBODIES, whereas for SHALLOW PITS and SHALLOW P-LAYERS only *neglubokij* can be used. Thus, *nevysokij* and *neglubokij* are more regular antonyms for *vysokij* and *glubokij* correspondingly; this is not surprising given that *nevysokij* and *neglubokij* are derived from *vysokij* and *glubokij*, respectively, by means of the negative prefix *ne-*. However, as Rakhilina (2000) argues, *nevysokij* and *neglubokij* are well-integrated lexicalized items in the Russian lexicon, rather than being compositional combinations of negation and their positive counterparts.

Let us now examine the cases of colexification (co-expression of the two frames by one and the same lexeme) that form the basis of our map. For the convenience of

discussion, we split our semantic map into the two parts: LATUS, which comprises the frames whose labels begin with THICK- and WIDE-, and ALTUS, to which belong the frames indicated by the labels beginning with LONG-, HIGH-, or DEEP- (and their mirror reflections in the antonymic zone of the "small" sizes in both cases). This division is represented by the dashed line on the maps above. The key reason for this split is that the cases of colexification within these two parts occur far more often than across the dashed borderline. We will address each of these parts in the subsequent sections, and also discuss the DEEP ~ THICK polysemy, which is the only bridge that connects the parts. However, the suggested split does not reflect any aprioristic theoretical position about the structure of the field.

3.   Patterns of lexicalization: LATUS

The most intriguing finding about the subzone of LATUS is that one kind of system seems to be the most frequent in our sample – the one exemplified by English, which discriminates only between *wide* and *thick* in the large size subdomain, and between *narrow* and *thin* in the small sizes subdomain. In other words, there is only one boundary, which cuts across both of the columns of the semantic map.

Such binary systems which consist of only two terms are very common in European languages, and are also consistently found outside of Europe. Hebrew, Udmurt, Armenian, and Ingush merge THICK LAYERS with THICK PIVOTS in using one term to cover this part of the domain, while also merging WIDE STRIPES with WIDE ROADS, TUBES, and HOLES which are covered by another lexical term. In the majority of cases, there is a symmetrical distinction in the antonymical field between 'thin' and 'narrow'.

ečkə	THICK I-LAYERS THICK PIVOTS	THIN I-LAYERS THIN PIVOTS	*ťejn'ə*
kel'i	WIDE STRIPES WIDE ROADS WIDE TUBES & HOLES BROAD SURFACES	NARROW STRIPES NARROW ROADS NARROW TUBES & HOLES	*šuvan'ə*
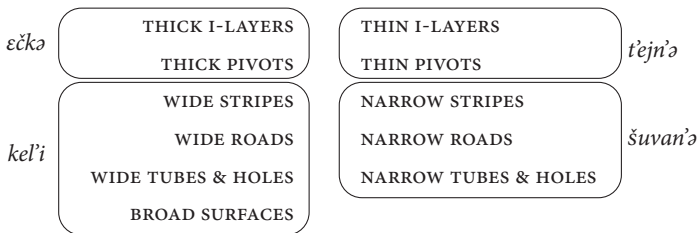
Figure 16. Semantic map of LATUS: Moksha

We suggest that this pattern of lexicalization be considered canonical. The concept of a canon in linguistic typology was introduced by Greville Corbett (see e.g., Brown et al. (eds) 2013); for Corbett, it means the set of properties of an "ideal" morphosyntactic construction that can serve as a point of departure for description

of typological diversity; here, we apply the term to a lexical system. Binary systems in the field of WIDE & THICK (and, later on, ternary systems in the case of LONG, HIGH, and DEEP) will serve as a point of departure, and we will describe cross-linguistic variation in these semantic fields by introducing additional lexical oppositions or eliminating the already proposed ones as we describe specific languages. Importantly, although what we identify as the canonical systems happen to be the most frequent in our dataset, yet we would like to emphasize that within this paper, *canonic* systems should not be equated cross-linguistically with 'the most frequent'. Rather, this term has to do with the fashion of describing the variation that we adopt: the observations made in terms of preservation or deletion of what we start out with as the canonical boundaries between frames can be insightful.

Binary systems drastically outnumber all the other systems in our sample. However, they are far from being universal. Almost half of the languages of our sample feature classifying systems which either introduce additional subtle lexical distinctions or, on the contrary, dispose of the boundary between THICK or THIN PIVOTS, and WIDE or NARROW STRIPES.

The second most frequent distinction is observed between THICK (or THIN) LAYERS and THICK (or THIN) PIVOTS: these two meanings are often expressed by different words. An example of a language which draws such a distinction is Kazakh, as in (4).

(4) KAZAKH
    a. *qalıŋ(*žuan) mata*
        thick   cloth
    b. *žuan (*qalıŋ) žip*
        thick   rope

(5) a. *žıka (*žiŋiške) žibek*
       thin   silk
    b. *žiŋiške (*žıka) tajaq*
       thin   stick

qalıŋ	THICK I-LAYERS	THIN I-LAYERS	*žıka*
žuan	THICK PIVOTS	THIN PIVOTS	*žiŋiške*
	WIDE STRIPES	NARROW STRIPES	*enziz*
qen	WIDE ROADS	NARROW ROADS	*tar*
	WIDE TUBES & HOLES	NARROW TUBES & HOLES	
	BROAD SURFACES		
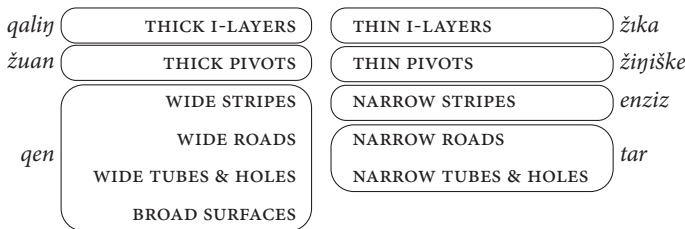
Figure 17. Semantic map of LATUS: Kazakh

It is sometimes claimed in Caucasian linguistics that the strategy of lexically distinguishing THICK LAYER from THICK PIVOT, and similarly distinguishing THIN LAYER from THIN PIVOT is a peculiar trait of the Caucasian linguistic area (Abaev 1995: 504; Klimov 1978:. 23–24; Chirikba 2008); the strategy is attested in Kartvelian (Georgian and Svan), North-West Caucasian (Abkhaz, Adyghe and Kabardian), and North-East Caucasian (Dargwa, Tsakhur) languages, as well as in Ossetic, the sole indigenous Indo-European language of this area. However, opposite to what is stated by Chirikba (2008), the THICK / THIN Layer vs. THICK / THIN Pivot distinctions are widespread outside the Caucasus: besides Kazakh, we find it, for instance, in Mandarin Chinese, Tatar, Tundra Nenets and Akebu (Kwa).

This distinction may not be drawn symmetrically in both of the subdomains of large and small size. In Tundra Nenets, the two adjectives for THIN LAYERS and PIVOTS (*n'anc'iko* and *yabta)* correspond to a single stative verb for THICK (*n'aŋoq-)*, which is applicable to both topological classes. The same can be observed in Northern Khanty, as well as in Lezgian and Aghul, which are Nakh-Daghestanian languages. The opposite strategy has not been attested in our sample; however, as Klimov (1978) reports, it is found in the other two languages of this family, Udi and Chamalal, where the single THIN-term corresponds to two distinct THICK-terms. (However, the data from the variety of Udi which is available to us contains only one THICK-term.)

The semantic subzone corresponding to the English lexeme *wide* (the WIDE-terms) seems to be less heterogeneous in terms of typological variety. This zone contains four frames which can be expressed either by means of a single lexeme (like in English) or by two different lexemes. In the latter case, the semantic space can be divided in three ways, as it contains three boundaries (WIDE Stripes vs. WIDE Roads vs. WIDE Tubes & Holes vs. BROAD Surfaces). All of the possible kinds of oppositions are attested in one or two languages within our sample.

An example of a specialized term for Surfaces is the Ižma Komi word *paš'kɨd*. According to the dictionaries of Standard Komi Zyrian, *paš'kɨd* (an adjectival derivate from the verb *paš'kədnɨ* 'expand, stretch out; widen', which in its turn is derived from the word *paš'* 'ajar') can describe the width of Stripes, Roads, Tubes, Holes and Surfaces, just as its cognate *paš'kət* in Udmurt. However, in Ižma Komi, which is the easternmost dialectal variety of Komi Zyrian for which we possess field data, the issues are somewhat more complicated: the width of Stripes, Roads, Tubes & Holes is described by another word, *ota* (a newer adjectival derivate from the word *ot* 'linen; expanse, width'). As for *paš'kɨd*, it is confined in Izhma Komi to wide Surfaces:

(6) IŽMA KOMI
    a. *paš'kɨd (*ota)   tɨ*
        wide   lake

b.   *ota (*paš'kɨd)   tuj*
     wide   road

A similar situation is found in Meadow Mari, where the word *lopka* can be used to describe Wide Stripes, Roads, and Tubes & Holes, but not Broad Surfaces for which a specialized word, *kumda*, is provided.

In Avar, we encounter a slightly different situation. Wide Roads, Tubes, and Holes can be described by two different adjectives: ʕebab and ʕat'idab. The first of them can also describe Wide Stripes, while the latter cannot. Conversely, ʕat'idab, unlike ʕebab, is applicable to Wide Surfaces. Avar thus presents a fairly rare (in our dataset) example of a system with merged Layers and Pivots in the subdomain of 'thick' accompanied by fine-grained distinctions in the subdomain of 'wide'. This is represented in Figure 18.
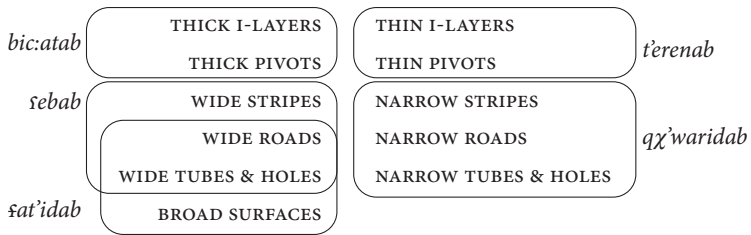
Figure 18. Semantic map latus: Avar

Another case for which the boundary between Stripes vs. Roads becomes relevant has been already described in Section 2.1: Kazakh has a specialized means of describing narrow Stripes as opposed to narrow Roads and narrow Tubes & Holes. Finally, Tundra Nenets manifests yet another, rare for our sample, type of distinction in this zone, the one between Roads, on the one hand, and Tubes & Holes, on the other. In the subdomain of small size, this language has the word *tiya* 'narrow' which can apply only to Stripes and Roads, and the word *pik* which refers only to Tubes & Holes.

Figure 19. Semantic map of latus: Tundra Nenets

Thus, we have discussed the two ways in which a language can deviate from the canon of the binary system in the subdomain of large size: it may establish further distinctions either in the subzone of the THICK-terms, or in the subzone of the WIDE-terms. The third possibility is to conflate across the boundary between the THICK-terms and the WIDE-terms, i.e., between THICK PIVOTS and WIDE STRIPES. That is what languages often do, and an important generalization to ensue from this fact is that such conflation occurs only within the domain of small size.

An illustrative example is the system that exists in Northern Khanty: it has a clear-cut binary system in the subdomain of large size, but in the antonymical small size subdomain the situation is more nuanced (Figure 20). As seen in the examples in (7), the word *uoχəł* can describe only THIN LAYERS, whereas another term, *vaś*, applies to THIN PIVOTS, STRIPES, ROADS, and TUBES & HOLES. Thus, Northern Khanty maintains the distinction between THIN LAYERS and THIN PIVOTS, which is rather unusual, and at the same time it eliminates the more widespread distinction between THIN PIVOTS and NARROW STRIPES which is supported in the majority of languages in our sample.

(7) NORTHERN KHANTY
a. *kuł soχ / uoχəł (*vaś) soχ*
    thick skin   thin   skin
b. *kuł juχ / vaś (*uoχəł) juχ*
    thick stick   thin   stick
c. *vutəŋ as / vaś (*uoχəł) as*
    wide river   narrow   river
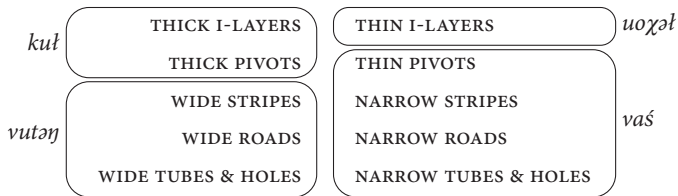
Figure 20. Semantic map of LATUS: Northern Khanty

A similar strategy is found in Kabardian (Figure 21), where, on the one hand, there is the usual NARROW-term *bʁʷəze* which can describe Stripes, Roads, Holes, and Tubes, and on the other hand, there is the word *psəʁʷe* which can apply to Pivots along with Stripes, Roads, Holes, and Tubes. (The word *psəʁʷe* is denotationally equivalent to *vaś* in Khanty.)
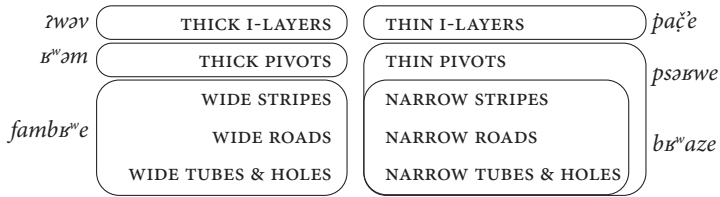
Figure 21. Semantic map of LATUS: Kabardian

Finally, another Caucasian language, Udi, appears to present a rather extreme case of the type of conflation described above (Figure 22). Udi has but one adjective for all of the discussed topological classes in the subzone of small size. The word *næzik'* conflates five distinct frames – THIN LAYERS, THIN PIVOTS, NARROW STRIPES, NARROW ROADS, and NARROW HOLES & TUBES.
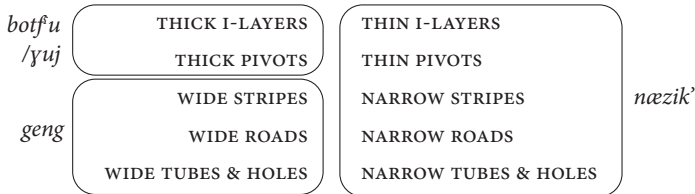
Figure 22. Semantic map of LATUS: Udi

To sum up, the typological variation in the domain of LATUS is mainly limited to the following options:

1. The binary strategy, which divides each of the large and small subzones into two, resulting in a pair of terms in each subzone ('wide' and 'thick' vs. 'narrow' and 'thin'), is the most frequent in our sample.
2. The second most frequent distinction is, in most cases, superimposed upon the binary system, when a boundary is being drawn – in both the small and large subdomains, or only in one of the two – between THICK / THIN LAYERS and THICK / THIN PIVOTS.
3. Further distinctions can be drawn in the subzone of 'wide', where there can be specialized terms exclusively applying only to some of the topological classes, i.e. to Surfaces or Stripes.
4. The boundary between THICK / THIN PIVOTS and WIDE / NARROW STRIPES, which occurs in the canonical binary systems, can be violated, yet only in the subzone of small size.

The extent to which the canonical binary systems are cross-linguistically recurrent requires a separate explanation. Trivial as they may seem for a speaker of a Standard Average European language, they in fact pose a challenge to semantic theory, for they are quite unexpected from the purely geometrical point of view. Why should languages describe such different objects as *cloth* or *paper*, and *sticks* or *ropes* in the same fashion? Why are *boards*, which usually have three distinct dimensions (length, width, and thickness), being merged with *holes in the fence,* for which width is the only dimension? The objects that are being united within such a lexical strategy bear little resemblance if any.

Perhaps the explanation should be sought in the typical ways that humans interact with, use, or observe (cf. Wierzbicka 2006) the objects described. Typically, an individual evaluates the width of a Hole when s/he looks at it, or puts her finger or her hand inside it, or even moves through it. Holes and Tubes are usually perceived from inside, and this quality relates them to Roads. The prototypical members of the Road class are roads, paths, bridges, and similar objects; most often people perceive their width from between their edges. As for Layers and Pivots, they are typically perceived from outside; in the most prototypical cases this is apparently done with the hand. Thus, the binary system presumably reflects the major modes of interaction with objects, which do not immediately follow from their shape.

To phrase it differently, the core idea of the concepts that correspond to terms like *wide* or *thick* may lie not in the field of the visual, but rather of the sensori-motor perception. In recent decades, a number of experimental studies have appeared corroborating the idea of embedded cognition (Varela et al. 1991; Lakoff & Johnson 1999). For example, it has been shown that interpretation of action words includes activation of the motor areas in the brain responsible for this kind of action (Pulvermüller 2005). The same may be true for dimensional terms: the concept of width, for example, may have to deal with our body schema: while being inside an object, we evaluate the distance from its edges to the confines of our body (this then explains why a desk is wide in the dimension which runs across the direction of our view, and not in the dimension which is along our line of sight: we are interested in how much space there is for our elbows.) The concept of thickness presumably may be grounded in tactual perception, for this is how we usually estimate the size of a piece of fabric, a rope, etc.

The fact that the merger of the boundary between THICK / THIN PIVOTS and WIDE / NARROW STRIPES occurs only in the subdomain of small sizes also seems to be explicable. Small size in general smooths out topological differences. For elongated objects like Stripes or Pivots one of the dimensions, i.e. their length, is considerably larger that the two others; basically, it is the difference between the two secondary dimensions that distinguishes Stripes from Pivots. For Stripes, the second dimension must be much wider than the third, while for Pivots the second and third

should be equal or almost equal. So if the second dimension of a Stripe shrinks, it becomes increasingly similar to a Pivot: a *narrow board* (a *strip of wood*) resembles a *thin stick* much more than *a wide board* resembles a *thick wooden log*. This may be the experiential motivation behind the systems we observed in Khanty and Kabardian.

4. On categorization of penetrable layers

Before we continue with the discussion of the ALTUS terms, we will have to focus on the specific linguistic behaviour of Penetrable Layers. *Snow*, *mud*, and *sand* exhibit a very peculiar topological behaviour in the languages of our sample. Unlike other Layers involving substances, the ways humans interact with *snow, mud,* and *sand* include wading through them, sinking into them, and being inside them (at least to a certain extent, for example knee-deep).

This is why in English and many other languages nominals that denote these substances can combine with terms like *deep* and *shallow* – of course, in the appropriate situations. For example, the substance should be lying on a road and not, say, on a roof (in the latter case they are described as being *thick*, similarly to Impenetrable Layers, cf. the discussion in Section 5.2.2). However, this pattern also turns out to be a parameter of typological variation.

In French, even the snow which lies around the porch and in which you are going to sink up to your waist is described with the same adjective that applies to THICK IMPENETRABLE LAYERS – *épais*. The antonymical adjective, *fin,* is used for the opposite meaning, as seen in (8).

(8) FRENCH
a. *tomber dans la   neige **épaisse*** / **profonde*
   fall   in   DET snow thick   deep
   'Fall into deep snow'
b. *marcher dans la   neige **fine*** / **peu profonde*
   go   in   DET snow thin   little deep
   'Step into not deep snow'

We find a similar pattern in two Turkic languages, Kazakh and Kirghiz, as well as in Lezgic (Nakh-Daghestatian). See, e.g., Figure 23 for Kazakh:
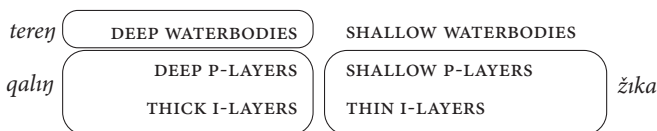
Figure 23. Colexification patterns of PENETRABLE LAYERS: Kazakh

In Irish the adjective *tanaí*, which is used primarily for ᴛʜɪɴ Lᴀʏᴇʀs (9a–b), is also the conventional way of expressing sʜᴀʟʟᴏw ᴘᴇɴᴇᴛʀᴀʙʟᴇ ʟᴀʏᴇʀs (Kukhto et al. 2016); meanwhile, the term *éadoimhin* which describes sʜᴀʟʟᴏw wᴀᴛᴇʀ ʙᴏᴅɪᴇs is infelicitous here.

(9) ɪʀɪsʜ
  a. *cuir   smearadh **tanaí** ime   ar   na   brioscaí*
      put.ɪᴘᴠ.2sɢ layer   thin   butter.ɢᴇɴ over ᴅᴇꜰ *cracker*.ᴘʟ
      'Spread a thin layer of butter over a cracker.'
  b. ⟨…⟩ *níl   ach   sneachta **tanaí** (*****éadoimhin) ar an   dtalamh.*
      ɴᴇɢ.ᴄᴏᴘ.ᴘʀᴇs ᴄᴏɴᴊ snow   thin   not.deep   on ᴅᴇꜰ ground
      '⟨…⟩ there is shallow snow on the ground.'

The situation with large size terms in Irish is quite the opposite: the thick-term *tiubh* is banished with Penetrable Layers, and only the deep term *doimhin* is possible, illustrated in (10). Figure 24 maps the Irish lexemes in this domain:

(10) ⟨…⟩ *thit sé insa   tsneachta doimhin (*****tiubh)*
      fall.ᴘᴛ 3sɢ in.ᴅᴇꜰ snow   deep   thick
      '⟨…⟩ he fell into the deep snow.'
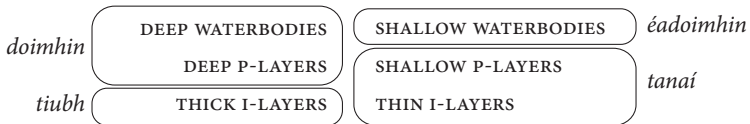
Figure 24. Colexification patterns of Pᴇɴᴇᴛʀᴀʙʟᴇ Lᴀʏᴇʀs: Irish

This pattern seen in Irish resembles the previously discussed Tʜɪᴄᴋ / Tʜɪɴ Pɪᴠᴏᴛs and Wɪᴅᴇ / Nᴀʀʀᴏw Sᴛʀɪᴘᴇs merger, which is only attested in small sizes. For human to be placed *within* a Penetrable Layer, it is indispensable that the Layer have some depth; otherwise, the human would conceive herself as standing *on* it. That is why when *snow* or *mud* is abundant, it is closer to the prototype of a Penetrable Layer, but when it is shallow, it rather resembles a plain Layer and is more likely to be described with ᴛʜɪɴ-terms rather than with sʜᴀʟʟᴏw-terms. However, Irish is the only language in our sample which exhibits this kind of asymmetry, so it is subject to further investigation.

5.   Patterns of lexicalization: ALTUS

5.1   Ternary systems

We are going on now to discuss the cross-linguistic behaviour of terms covering the frames of the ALTUS subzone. In most European languages in our sample, the semantic domain discussed in this section is split into three sections: first, height; second, depth; and third, length. Thus, as in the case of the LATUS terms we have just discussed, there exists a canonical way to divide the conceptual space; the only difference is that the division is ternary rather than binary, so we expect to find canonical systems with three terms for small size and three terms for large size. Indeed, such systems are attested in more than half of the languages in our sample. Some of them belong to the Indo-European language family (Armenian, Serbian, Spanish), but there are languages from other genetic groups and linguistic areas as well (Azeri, Kabardian, Georgian, Ingush). English does not manifest a canonical ternary system if only because it draws an additional distinction between *high* and *tall*, which we address in more detail below.

Let us look closer at the system of dimensional terms in Besleney Kabardian (Northwest Caucasian). It consists of exactly six lexemes: *kʷwə* 'deep' and *čʼenǯʼ* 'shallow' which describe the depth of Pits and Penetrable Layers; *λaye* 'high/tall' and *λaχčʼe* 'low' for Poles and Barriers; and finally, *čʼəh* 'long' and *čʼagʷe* 'short' for the length of Pivots.

čʼəh	LONG PIVOTS	SHORT PIVOTS	*čʼagʷe*
λaye	TALL POLES / HIGH BARRIERS	SHORT POLES / LOW BARRIERS	*λaχčʼe*
kʷwə	DEEP PITS / DEEP WATERBODIES / DEEP P-LAYERS	SHALLOW PITS / SHALLOW WATERBODIES / SHALLOW P-LAYERS	*čʼenǯʼ*
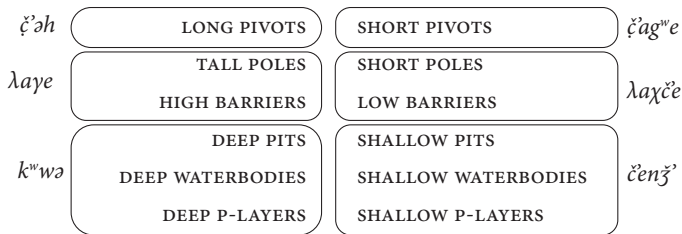
Figure 25. Semantic map of ALTUS: Besleney Kabardian

This is a perfect, clear example of a canonical system. Then, dimension terms in many other languages can be described as deviating somewhat from such a system if they do not follow the ternary logic so neatly. Such languages retain the division into these three domains, but introduce additional distinctions, so that the systems become richer.

The domain of height in English is split between two adjectives, *high* and *tall*. On the basis of a corpus study, Dirven and Taylor (1988) put forward several parameters that determine the usage of *tall*, namely, extreme prominence of the

vertical dimension, and thin and slender shape. They explain the high frequency of collocations of *tall* with animate beings, postulating a 'dynamic component' in its meaning. Indeed, according to the corpus[3] data, *tall* most frequently occurs with nouns denoting people and trees. However, if we take into account the general frequency of particular noun lexemes and look at the relative frequency scores of the combinations of *tall* with different names of objects, it turns out that in the British National Corpus, words which co-occur with *tall* with the highest relative frequency are *chimney* and *mast*. Collocations with *stool*, *grass*, *tower*, *building* and *tree* immediately follow them, and human beings come only afterwards. It is clear that the differential property of these objects is not "dynamicity" in any sense, but their elongated form and fixed (or at least prototypical) vertical orientation, so in our terms they belong to the same topological class of Poles.[4] Thus, the domain of height corresponds to two cross-linguistically distinct functions (=frames, in our talk), because there are several languages that encode the height of Barriers and the height of Poles in different ways.

There is one more difference between *tall* and *high* in English, and it concerns their metonymical uses. Only *high*, but not *tall*, can describe the distance of an object which is above the ground, to the ground, e.g. *high /[??]tall ceiling*. The existence of such metonymical uses is also a parameter of variation among languages. For instance, in Izhma Komi (Permic, Finno-Ugric), dimensional terms do not exhibit this kind of metonymy.

If we look at the domain of depth, we will find that it is not homogenous either. A good example of a split can be taken from Russian. Russian has two adjectives to describe small depth: *neglubok'ij* and *melk'ij*. The first is a mirror reflection of the word *glubokij* 'deep', and covers SHALLOW PITS, SHALLOW WATERBODIES, and SHALLOW PENETRABLE LAYERS. The second adjective in this domain combines only with WATERBODIES. The adjective *melk'ij* is etymologically derived from the noun *mel'* 'shoal'.[5]

---

3. Our English data are collected by means of BYU-BNC (Brigham Young University – British National Corpus, http://corpus.byu.edu/bnc/), COCA (Corpus of Contemporary American English, http://corpus.byu.edu/coca/), and Google.

4. R. Dirven and J. Taylor (1988) define two meanings of *tall*. The first one is described above, and the second one is "massive, bulky objects, that silhouette against background" which aims to explain its combination with nouns like *mountain* and *pyramid*. However, such collocations are quite rare in the corpora and we suppose that mountains and pyramids described as *tall* are merely coerced into the Pole conceptual template: they are just non-prototypical Poles.

5. *Melkij* can be used with Pits as well, but such cases are probably a manifestation of another meaning of this adjective which does not denote the depth, but instead characterizes the general size of a pit or a ditch as a whole. This meaning does not confine itself to Waterbodies, cf. *melkij žemčug* 'small pearls', *melkaja ryba* 'small fish'.

(11) RUSSIAN
  a. *neglubokaja reka melkaja reka*
     shallow   river shallow river
  b. *neglubokaja jama* #*melkaja jama*
     shallow   pit   small   pit'

Similar terms which can only denote SHALLOW WATER BODIES, but not SHALLOW PITS, are found in Moksha Mordvin and in Chukchi.

According to our data, no splits are ever added to the domain of 'long'. In the languages of our sample, there is either one pair of terms for the frames of LONG PIVOTS and SHORT PIVOTS, or several terms which describe length or shortness on an equal footing, and differ from each other in their ability to be used in some other contexts (for example, to describe LONG POLES or SHORT POLES).

To summarize, one of the ways to deviate from a canonical system is to introduce further parameters. As a result, a language which exhibits such deviation possesses more than three pairs of lexemes in the domain of 'long', 'high', and 'deep'.

5.2 Unifying systems

As we have shown in the discussion of 'wide' and 'thick' in Section 3 above, one way in which lexical coverage of a domain can depart from a canonical system is to merge or unify the expressions of a subdomain. A unifying system is the term adopted in MLexT for systems that merge more than a canonical system does (also see Chapter 1). In this case, at least one of the boundaries on the canonical map disappears, and the two domains amalgamate, being expressed by a single term. We distinguish between two types of unifying systems, as each of the two boundaries – the one between depth and height (more specifically, between DEEP PITS and HIGH BARRIERS), and the one separating height from length (i.e. TALL POLES from LONG PIVOTS) – can be erased.

The first type of unifying strategy deals with the merger of height and depth; we shall call the systems that come about this way DH-systems. They seem to be especially frequent in Finno-Ugric languages, and this is probably not by chance. Five of the eight Finno-Ugric languages in our sample demonstrate the DH-system: Beserman Udmurt, Finnish, Izhma Komi, Moksha Mordvin and Northern Khanty. DH-merger is also attested at least in Classical Latin, as we have seen in the introductory section. The feature that all the DH-languages have in common is that they merge some topological classes from the domain of depth and some from the domain of height; however, they are rarely identical to each other: the exact number of topological classes involved in the merger varies across languages in our sample.

The second type of unification is based on the merger between the domains of height and length, which we will refer to as HL-merger. Among the seven Northeast

Caucasian languages of our sample, three (Aghul, Itsari Dargwa, and Avar) exhibit the HL-strategy. A similar pattern is found in two Turkic languages (Kazakh and Turkish) and one Finno-Ugric language (Izhma Komi). Izhma Komi exhibits both HL- and the DH-mergers; there are only two pairs of terms for Poles, Pivots, Barriers, Pits, Water Bodies and Penetrable Layers, and the boundaries between them in each subdomain do not coincide with either of the boundaries in the canonical system. The case of Izhma Komi will be discussed in detail in Section 5.2.2.

5.2.1   *DH-merger*

As pointed out in Section 1 above, DH-systems occur in Latin and at least in five (and probably more) Finno-Ugric languages. Nearly all the DH-systems are asymmetrical: as we saw in the case of LATUS terms (Section 3), adjectives of the large and the small size subdomains can combine with different numbers of topological classes. Northern Khanty and Finnish show the same type of asymmetry: they have a ternary system in the zone of large sizes, and they merge in the zone of small sizes. In Northern Khanty, the adjective *măł* 'deep' describes Penetrable Layers, Waterbodies and Pits, while *kărs'* 'high' occurs with Barriers and Poles. The adjective *łeł* 'shallow/low' covers the topological classes of Waterbodies, Pits, Barriers and Poles but is prohibited with Penetrable Layers. Objects like snow or mud cannot be described with any dimensional term of the small size subdomain at all. The only possibility for describing their size is to use the adjective *aj* 'small'; see (12).

(12)   NORTHERN KHANTY
    a.   *măł / łeł   as*
        deep   shallow river
    b.   *kărs' / łeł juχ*
        high   low tree
    c.   *măł / *łeł / aj   łɔn's'*
        deep   shallow   small snow

The Finnish system is similar, as exemplified in (13). The adjective *syvä* 'deep' describes P-Layers, Waterbodies and Pits, and *korkea* 'high' applies to Barriers and Poles. In the subdomain of small size, the adjective *matala* 'shallow/low' is used with names of all objects belonging to these topological classes; but just as in the case of Northern Khanty, it is rather infelicitous with Penetrable Layers.

(13)   FINNISH
    a.   *syvä / matala kuoppa*
        deep   shallow pit
    b.   *korkea / matala muuri*
        high   low   wall
    c.   *syvä / ??matala lumi*
        deep   shallow snow

The data on Finnish and Northern Khanty are summarized in Figures 26 and 27.

χuv | LONG STRIPES & PIVOTS | SHORT STRIPES & PIVOTS | *van*
kărs' | TALL POLES / HIGH BARRIERS | SHORT POLES / LOW BARRIERS | *łeł*
măł | DEEP PITS / DEEP WATERBODIES / DEEP P-LAYERS | SHALLOW PITS / SHALLOW WATERBODIES / SHALLOW P-LAYERS | *aj* 'small'
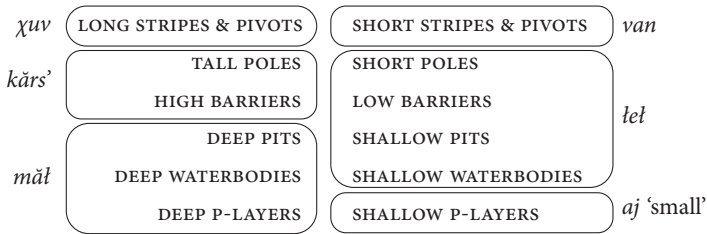
Figure 26. Semantic map of ALTUS: Northern Khanty

pitkä | LONG STRIPES & PIVOTS | SHORT STRIPES & PIVOTS | *lyhyt*
korkea | TALL POLES / HIGH BARRIERS | SHORT POLES / LOW BARRIERS | *matala*
syvä | DEEP PITS / DEEP WATERBODIES / DEEP P-LAYERS | SHALLOW PITS / SHALLOW WATERBODIES / SHALLOW P-LAYERS

Figure 27. Semantic map of ALTUS: Finnish[6]

A different DH-system is attested in Beserman Udmurt. The adjective *mur* 'deep' combines with DEEP WATERBODIES, DEEP PITS, and DEEP P-LAYERS. Pits, at the same time, may be described by another adjective *ž'užət* 'high/deep' that is also used with Barriers and Poles. Thus, although there are two terms for 'high' and 'deep', their distribution is not similar to the canonical one. The small size subdomain in Beserman Udmurt contains only one adjective *lapeg* 'low/shallow' that covers all the topological classes from the depth and the height domains except Layers. As in the case of Northern Khanty, an adjective of common size – *piči* 'small' – is used to describe it.

---

6.   Although in the majority of the languages in our sample people pattern with Poles, they may combine with a different set of dimensional terms; that is, people always constitute a separate and important micro-frame (see Chapter 1 in this volume). In Finnish, 'a tall man' is *pitka mies* (lit. 'a long man') rather than *korkea mies*. However, we do not reflect this fact in Figure 27. We are inclined rather to describe this issue as stemming from different categorizations of people as Pivots in some languages or Poles in others, than to introduce a separate frame into the map. Height of people can be described by general, non-dimensional size terms (cf. French *grand homme* 'tall person', where *grand* 'big' and not *haut* 'high' is used), but our dataset contains no examples of languages which would have a specialized term just for people.

(14) BESERMAN UDMURT
    a.   *ǯ'užət* / *mur* / *lapeg*   *gu*
         deep   deep   shallow pit
    b. **ǯ'užət* / *mur* / *lapeg*   *reka*
         deep   deep   shallow river
    c.   *ǯ'užət* / *lapeg ken'er*
         high   low   fence
    d. *mur* / *??lapeg* / *piči*   *lômâ*
         deep   shallow   small snow

Moksha Mordvin also exhibits a DH-system. The adjective *aln'ɛ* 'low/shallow' covers Pits, Barriers and Poles; some of the native speakers allow it with Waterbodies too, but it is fully unacceptable with Penetrable Layers, as exemplified in (15).

(15) MOKSHA MORDVIN
    aln'ɛ ber'ɛf / *šuftə* / *lotka* / *??l'ɛj* / **lov*
    low fence   tree   pit   river   snow

Unlike other Finno-Ugric languages, Moksha has the term *mac'ɛ* 'shallow' that entirely parallels the Russian *melkij*. The Moksha term can apply only to Waterbodies as in (16), and is probably historically connected to the word meaning 'shoal'.

(16) *mac'ɛ* *l'ɛj*   / **lotka*
    shallow river   pit

In the subdomain of large size Moksha also exhibits a DH-merger. The adjective *ser'i* 'deep/high' combines with Pits, Barriers, and Poles, and is even judged as acceptable with Waterbodies by some native speakers. The adjective *kər̥ka* 'deep' covers all the topological classes within the depth domain. Example (17) illustrates these collocations, and Figures 28 and 29 present the semantic maps for Beserman Udmurt and Moksha Mordvin, respectively.

(17) a. *ser'i šuftə* / *ber'ɛf* / *lotka* / *ok/?l'ɛj*
        high tree   fence   pit   river
    b. *kər̥ka lov* / *l'ɛj* / *lotka*
        deep   snow   river   pit

A curious observation can be made about the Finno-Ugric DH-systems. Among the words which exhibit the DH-merger (Khanty *łeł*, Finnish *matala*, Udmurt *lapeg*, Komi *l'apkɨd,* Moksha *aln'ɛ* 'low, shallow' and Udmurt *ǯ'užət*, Komi *džudžɨd*, Moksha *kər̥ka* 'deep, high') only two pairs (the Udmurt and the Komi terms *lapeg* and *l'apkɨd,* and *ǯ'užət* and *džudžɨd*) are cognates. This is not surprising, as Udmurt and Komi form the Permic group, which is a separate genetic unit of the Finno-Ugric family. However, the DH-pattern itself is not restricted to Permic

kuž' | LONG STRIPES & PIVOTS | SHORT STRIPES & PIVOTS | *vakči*

ž'užət

TALL POLES — SHORT POLES
HIGH BARRIERS — LOW BARRIERS — *lapeg*
DEEP PITS — SHALLOW PITS

mur — DEEP WATERBODIES — SHALLOW WATERBODIES
DEEP P-LAYERS — SHALLOW P-LAYERS — *piči* 'small'
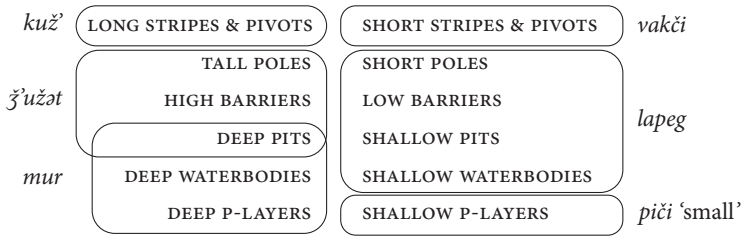
Figure 28. Semantic map of ALTUS: Beserman Udmurt

kuvaka | LONG STRIPES & PIVOTS | SHORT STRIPES & PIVOTS | *n'uṛken'ɛ*

ser'i

TALL POLES — SHORT POLES
HIGH BARRIERS — LOW BARRIERS — *aln'ɛ*
DEEP PITS — SHALLOW PITS

kəṛka — DEEP WATERBODIES — SHALLOW WATERBODIES — *mac'ɛ*
DEEP P-LAYERS — SHALLOW P-LAYERS — *jomla* 'small'
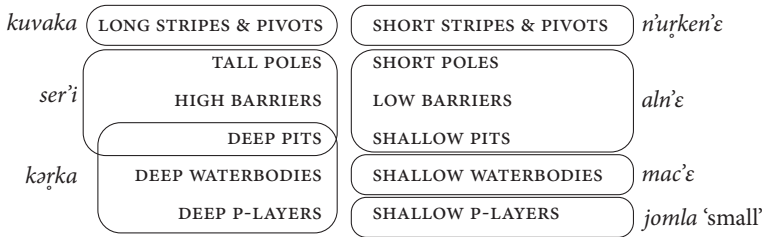
Figure 29. Semantic map of ALTUS: Moksha Mordvin

languages. We can hypothesize that it was already present in Proto-Finno-Ugric, and then was retained in different branches of the family in spite of the new lexical material entering the semantic field. Another interesting fact which seems to support this hypothesis is that the Moksha word *kəṛka* 'deep', which does not exhibit the DH-merger, is a cognate of the Finnish term *korkea* 'high, tall', which synchronically also has nothing to do with a merger (see Redei 1986). This can be explained if we assume that the ancestor of the two lexemes denoted both 'high' and 'deep', as is the case in numerous modern Finno-Ugric languages discussed above.

5.2.2 *HL-merger*

There are seven Northeast Caucasian languages in our sample: Aghul, Avar, Ingush, Itsari Dargwa, Lezgian, Tabasaran, and Udi. They are certainly not remotely representative of the whole Northeast Caucasian family, let alone of the whole "Caucasian linguistic area". We cannot state whether the pattern observed in this section is typical for the area or the family. HL-mergers are attested in three of our languages: Aghul, Itsari Dargwa, and Avar.

Only in Avar do the canonical and non-canonical strategies coexist. Lexemes for the large sizes follow the canonical pattern: the adjective *borxatab* combines with HIGH BARRIERS and TALL POLES, which are topological classes from the domain of height. Another adjective, *xalatab*, is used to describe LONG PIVOTS; compare the examples in (18).

(18)   AVAR
    a.   *borxatab qed*
        high   wall
    b.   *borxatab ʕvet'*
        high   tree
    c.   *xalatab kvar*
        long   rope

It is in the zone of small size that the merger arises. In this case, one and the same lexeme covers SHORT PIVOTS and SHORT POLES (thus violating the canonical boundary between these frames), and another lexeme specializes in LOW BARRIERS. Example (19) illustrates this pattern, and Figure 30 summarizes the system.

(19)   a.   *qχ'oqχ'ab kvar*
        short   rope
    b.   *qχ'oqχ'ab ʕvet'*
        low   tree
    c.   *t'ucab cali*
        low   fence
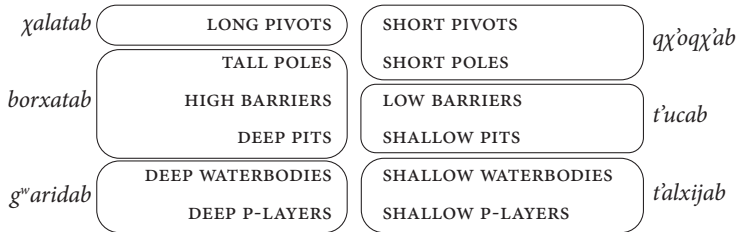
Figure 30. Semantic map of ALTUS: Avar

The other two Northeast Caucasian languages that exhibit HL-merger (Aghul and Itsari Dargwa) have identical systems of dimensional terms where large and small sizes are covered by two adjectives each. In the large size subdomain, the boundary runs between TALL POLES and HIGH BARRIERS, and in the small size subdomain the division is marked between LOW BARRIERS and SHALLOW PITS, thus completely merging the domains of height and length as opposed to the domain of depth; see Figures 31 and 32.

    Another, more consistent type of HL-merger is attested in Kazakh. In the zone of ALTUS, there are two terms both for large and for small sizes, and in both cases their scopes of usage intersect. HIGH or LOW BARRIERS, and TALL or SHORT POLES can be described by the same adjectives: *biik* 'high' and *alasa* 'low' apply both to fences and trees. Another pair, *üzın* 'tall/long' and *qısqa* 'low/short' describes Poles

irχe

	LONG PIVOTS	SHORT PIVOTS	
	TALL POLES	SHORT POLES	*dʒiqːe*
	HIGH BARRIERS	LOW BARRIERS	
hawa	DEEP PITS	SHALLOW PITS	
	DEEP WATERBODIES	SHALLOW WATERBODIES	*xeba*
	DEEP P-LAYERS	SHALLOW P-LAYERS	
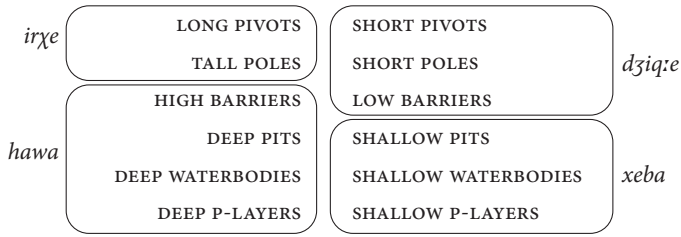
Figure 31. Semantic map of ALTUS: Aghul

buqen

	LONG PIVOTS	SHORT PIVOTS	
	TALL POLES	SHORT POLES	*qːantʼ*
	HIGH BARRIERS	LOW BARRIERS	
čaq	DEEP PITS	SHALLOW PITS	
	DEEP WATERBODIES	SHALLOW WATERBODIES	*ʁuk*
	DEEP P-LAYERS	SHALLOW P-LAYERS	
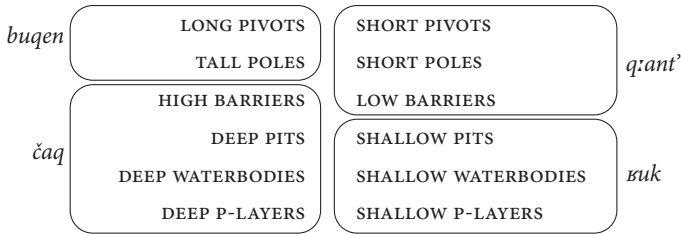
Figure 32. Semantic map of ALTUS: Itsari Dargwa

and Pivots – trees, sticks and ropes. Thus, Poles may be described by either pair of the adjectives, that is, the one which profiles topological similarities of Poles to Barriers, or the one foregrounding the fact that Poles resemble Pivots; see the examples in (20) and the semantic map in Figure 33.

(20)   KAZAKH
    a. *bïk   dual / ağaš*
        high   wall   tree
    b. *üzın ağaš / žip*
        long   tree   rope
    c. *alasa dual / ağaš*
        low   wall   tree
    d. *qısqa ağaš / žip*
        short tree   rope

üzın

	LONG PIVOTS	SHORT PIVOTS	*qısqa*
	TALL POLES	SHORT POLES	
bïk	HIGH BARRIERS	LOW BARRIERS	*alasa*
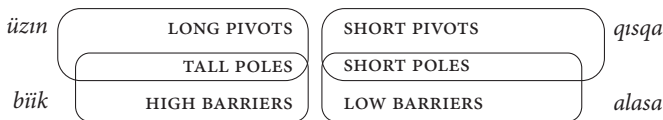
Figure 33. Semantic map of ALTUS: Kazakh

The most intriguing case, however, is presented by Izhma Komi, the discussion of which we have postponed until this very moment. The Komi system contains only four terms, but their distribution reflects both the DH and the HL strategies at the same time, thus maximally deviating from the canonical map.

On the one hand, the DH-merger is there, for large sizes as well as for small ones. Both the lexemes *džudžɨd* 'deep/high/tall' and *l'apkɨd* 'shallow/low' are used with Layers, Waterbodies, Pits, and Barriers. In addition, *l'apkɨd* even applies to Poles, but *džudžɨd* is generally judged unacceptable by speakers in this context; thus, the system is asymmetrical. On the other hand, the HL-merger also contributes to the configuration of the system. The main adjective describing HIGH POLES is *kuz'* 'tall/long', which is equally applicable to LONG PIVOTS. SHORT PIVOTS are described by the word *džen'ɨd* 'short', but it never describes SHORT POLES (unless they lose their vertical orientation and virtually become Pivots). The combination of these two strategies results in a highly asymmetrical system: in both the large and the small subdomain, the conceptual space is divided into two parts, and in both cases the place of the boundary does not coincide with the canonical one (21). The Izhma Komi system is illustrated in Figure 34:

(21)   IZHMA KOMI
    a.   *džudžɨd lɨm   / ju   / jama / zabor / ??pu*
       high   snow river pit   fence   tree
       'deep snow' / 'deep snow' / 'deep pit' / 'high fence' / 'tall tree'
    b.   *kuz' pu / gez*
       long tree   rope
       'long tree' / 'long rope'
    c.   *l'apkɨd   lɨm   / ju   / jama / zabor / pu*
       shallow_low snow river pit   fence   tree
       'shallow snow' / 'shallow river' /'shallow pit' /'low fence' / 'low tree'
    d.   *džen'ɨd *pu / gez*
       short   tree   rope
       'low tree' / 'short rope'

kuz'	LONG PIVOTS	SHORT PIVOTS	*džen'ɨd*
	TALL POLES	SHORT POLES	
	HIGH BARRIERS	LOW BARRIERS	
džudžɨd	DEEP PITS	SHALLOW PITS	*l'apkɨd*
	DEEP WATERBODIES	SHALLOW WATERBODIES	
	DEEP P-LAYERS	SHALLOW P-LAYERS	
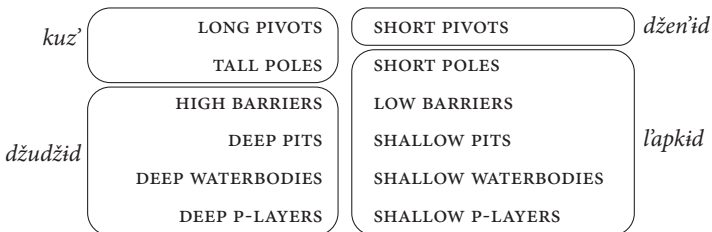
Figure 34. Semantic map of ALTUS: Izhma Komi

6.   Discussion

6.1   General remarks

We have now examined several semantic domains within the overarching domain of dimension, noting regularities and tendencies which govern lexification of dimensional meanings.

Attribution of an object to a certain topological class is based upon multifold information about this object; it considers not only shape, size, and orientation, but also the conventional patterns in which humans interact with the object, and furthermore, its functions in human life. When a language merges two frames into the denotation of a single lexeme, the merger picks up a subset of these properties. There are mergers motivated by the common form of objects from two classes (for example, TALL POLES and LONG PIVOTS, i.e. the HL-merge) and there are mergers motivated by similar sensorimotor impressions which arise from interaction with objects (THICK LAYERS and THICK PIVOTS, which are prototypically measured with the hand). Thus, the conceptual space is divided between several lexical items, resulting in a system of adjectives. Each system we have studied so far conforms to the generalizations we have proposed, formalized by means of our semantic map.

The regularities which govern lexification are fairly different across the subparts of the conceptual field of dimensions, as in each case they emerge from the specific topological and functional properties of the objects in question. Still, we can draw some further generalizations.

First of all, we have encountered systems which are canonical, that is, they correspond to our point of departure for describing typological diversity in this domain. These are ternary systems for 'long' vs. 'high' vs. 'deep', and binary systems for 'thick' vs. 'wide'. On the basis of such canonical systems, other systems have been described either as "unifying" if they dispose of one or more boundaries set by the canonical systems, or as "classifying" if they introduce further distinctions. In our sample, ternary 'long-high-deep' and binary 'thick-wide' systems are the most frequent, which may be partly influenced by a certain European bias of our sample. However, we have shown that such systems are also attested outside Europe.

Nearly all non-canonical systems exhibit asymmetry between the subdomains of large versus small size. In the unifying systems, asymmetry is almost obligatory. Usually, if in some language one of the subdomains merges more frames than the canonical system does, the other subdomain never follows it. Recall that none of the DH-languages discussed in Section 5.2.1 have a symmetrical system, and among the examined HL-languages the only symmetrical system is found in Kazakh. Similarly, the boundary imposed by the binary system for the 'wide'&'thick' zone can only be violated in the subdomain of small size, and it necessarily leads to asymmetry.

However, evidence of asymmetry between antonymical domains runs as a refrain throughout the studies presented in this volume, so it is not surprising at all that dimensional terms for large and small size are not mirror reflections of each other. What *is* in fact surprising is the multitude of symmetrical systems in the domain of dimensions. In Chapter 2, Kyuseva et al. argue that the conceptual fields of SHARP and BLUNT are structured in different ways, and so are YOUNG and OLD, and many other semantic domains studied in the projects of MLexT.

Dimensions, on the contrary, have at least two antonymic parts in their semantic map that are symmetrical to each other (aside from the single frame of WIDE SURFACES that has no antonymic counterpart). This in principle creates the logical possibility for existence of symmetrical systems, and this possibility is often realized. In fact, as compared to other domains reported on in this volume, dimensions are much more frequently lexicalized symmetrically. Within the canonical systems, the lexemes sometimes resemble a paradigm, and they are sometimes subject to paradigmatic effect. According to (Zaliznyak 1985), in Late Common Slavic, dimensional terms used to belong to the *ŭ*-declension; later, along with all other adjectives of this declension, they migrated to the adjectival declension, attaching the "empty" adjectival suffixes *-ъk-* or *-ok-*. Earliest manuscripts display a great deal of variation, but later all the dimensional terms levelled according to the same pattern: the small size adjectives invariably use *-ъk-* (*korot-ъk-ъ* 'short', *niz-ъk-ъ* 'low', *měl-ъk-ъ* 'shallow', *tъn-ъk-ъ* 'thin', *ǫz-ъk-ъ* 'narrow'), while their large size counterparts contain only *-ok-* (*vys-ok-ъ* 'high', *glǫb-ok-ъ* 'high', *šir-ok-ъ* 'wide'). (For further typological perspectives with respect to "canonical" antonymical pairs, see Jones & al. 2012; Paradis & al. 2009; Koptjevskaja-Tamm & Miestamo 2015; cf. also Murphy 2003.)

We can observe some additional regularities governing the possible degree to which systems can be asymmetrical. These regularities concern asymmetries which emerge from elimination of the boundaries established in the canonical systems, i.e. those found in unifying systems. In some of them, a merger affects only the subdomain of small size, whereas that of large size presents a perfect canonical system. In other languages, merger is found in both of the subdomains, but never does it occur in the large size subdomain without affecting the small size subdomain. On the basis of this observation, we formulate the Dimensions Merger Constraint on possible unifying systems.

> The dimensions merger constraint
>
> If a language eliminates the canonical boundary in the subdomain of large size, it must also dispose of the corresponding boundary in the subdomain of small size.

This constraint works as an implicational universal in the Greenbergian sense (Greenberg 1966). That is, it predicts that there are no languages with mergers over the canonical boundaries in the large size zone that preserve boundaries within the small size zone at the same time. We hypothesize that the constraint is due to the fact that boundaries between small objects are more flexible, as small objects are perceptually more similar to each other than large objects. As we have already stated, the topological differences across small objects are also less pronounced, and therefore less important for categorization.

Our last observation concerns Penetrable Layers, which include items like snow and mud. Penetrable Layers constitute a boundary between ALTUS and LATUS, being themselves a somewhat marginal class. Some languages (e.g. Khanty, Finnish, and Chukchi) do not describe Penetrable Layers with dimensional terms (especially with the terms referring to small sizes) at all. When the depth of Penetrable Layers is lexicalized, the depth expression can be colexified either with thickness of Impenetrable Layers or with depth of Waterbodies. However, there is a strong tendency for lexemes which apply to these topological classes to not extend over the class of Penetrable Layers if they participate in other mergers. If there is a DH-merger in a language (as in several Finno-Ugric languages), then snow and mud are very unlikely to be described as deep; if a language colexifies THICK PIVOTS and THICK IMPENETRABLE LAYERS (as most European languages do), then we again should not expect it to apply the same lexeme to PENETRABLE LAYERS.

The last fact is quite understandable if we return to the claim made in the beginning of this section: mergers of adjacent frames occur because the objects corresponding to these frames are similar from the point of view of human experiencers. This is why lexicalization should be, at least to a certain degree, consistent. We should not expect a language to add DEEP PENETRABLE LAYERS to the cluster which already consists of THICK PIVOTS and THICK IMPENETRABLE LAYERS, as the latter merger has a sensorimotor motivation, and what ties Penetrable and Impenetrable Layers together is their visual similarity, i.e. the geometrical form.

6.2   M. Bierwisch and E. Lang: A componential analysis approach

Unlike several other semantic domains under consideration in this volume, dimensional terms have previously been subject to thorough semantic description. The most prominent approach is that suggested by M. Bierwisch and E. Lang (1989) who developed a rather complicated semantic theory within the framework of componential analysis.

According to M. Bierwisch and E. Lang, the semantic representation of a dimensional term consists of a bunch of primes. Firstly, the perceptual system must

pick out one, two, or three axes in the object under consideration (thus coercing it into one of the 26 object schemata which are very similar to what we call topological classes). Then spatial primes (Max, Min, Vert, Across, and Obs) can be assigned to these axes according to their spatial characteristics. Finally, subsets of such primes are lexicalized into specific dimensional terms. One prime can specify another, e.g. Max (= the maximum axis without additional specification) stands for the English lexeme *long*, and Max | Vert (=the maximum axis which is at the same time vertically oriented) stands for *tall*.

Typological variation in the domain of dimensions is basically treated by M. Bierwisch and E. Lang as differences in categorization. First of all, in their system the variation is concerned with the choice of the axis to which the Across prime is ascribed. The so-called ACROSS-terms are the lexemes corresponding to English *wide*. The ACROSS label captures the generalization that the dimension which is described as *wide* always goes *across* some more salient axis: typically, it is the largest dimension (as in the case of a *wide board)*, or the direction of the Observer's look *(wide desk)* that is considered the most salient.

However, not all languages allow variation in the possible categorization of a single object.E. Lang (2001) reports the results of psycholinguistic experiments conducted with speakers of Mandarin Chinese and Korean. The participants were shown a wooden parallelepiped in different settings (such as a table top, a sideboard, a chest of drawers, and a window-sill) and were asked to describe it with dimensional terms. The Mandarin Chinese speakers consistently used the 'long'-term *cháng* for the largest dimension, and the 'wide'-term *huān* for the second largest one, no matter how the parallelepiped was situated with respect to the Observer. Korean speakers employed two strategies. The first one was to describe the wooden board with a pair of terms *kili* and *phok*, which were used according to the Chinese pattern. The second strategy was to use two other terms, the meaning of which strongly depended on the axis of the Observer's view: the 'long'-term *selo* always coincides with the axis of view, while the 'wide'-term *kalo* always designates the dimension that goes across the axis of view. Thus, Chinese and Korean speakers would differ in description of, say, a sideboard: its largest dimension (which typically goes across the axis of the view of the human who is going to take something from it or put something into it) was described with the 'long' term by the Chinese participants, but with the 'wide' term by the Korean participants.

Consequently, M. Bierwisch and E. Lang divide languages into three types which they refer to as P-fixed, O-fixed and P/O mixed, according to the dimension that a language can designate as 'wide': the one going across the maximum dimension (like in Chinese), the one which is athwart the Observer's direction of view (like in Korean), or both (like in English). It turns out that in the P/O mixed languages, one more opportunity is sometimes possible: the dimension which crosses

the upright vertical axis can also be 'wide' (although there is no language for which applying 'wide' to this axis is the only possibility).

In our view, these results, interesting as they are, do not reflect the most substantial part of typological variation in the domain of dimensions, namely, how different frames are lexicalized in various ways. The model of Bierwisch and Lang overlooks such important discrepancies between languages as the oppositions in the subzone of 'wide' (the major European pattern vs. Avar or Tundra Nenets), or the merger of 'thick' and 'deep', as in Irish. While their model embodies certain significant insights into the semantic representation of dimensional terms in many European systems, it definitely lacks empirical coverage when applied to a wider sample of languages.

At the same time, the aforementioned difference between Mandarin Chinese and Korean represents another dimension of typological variation. In our terms, what Bierwisch and Lang attempted to capture is cross-linguistic difference in categorization of an object as a member of a certain topological class. At this point, let us recall that a cupboard (or, generally speaking, a wooden parallelepiped) is in fact a rather non-prototypical object to be described with dimensional terms. We suggest that a cupboard, if described as *long* and *wide*, is in fact a fairly non-prototypical Stripe – because in our system a physical object (as opposed to an empty space) that can be described as 'wide' can only be a Stripe. Note that in a typical situation when the dimensions of a cupboard are discussed, only two of the three dimensions of a cupboard are relevant (e.g. when the goal of the conversation is to figure out how to fit a cupboard in a particular corner of a room). So the difference between Chinese and Korean is the exact way those languages squeeze such concept as cupboard into the category of Stripes.

The core concepts of our topological classes are to some extent culturally independent, such as natural objects like trees, waterbodies, and roads, or they are objects which presumably are used in identical ways in different cultures such as ropes, sticks, cloth, etc. Whether pieces of furniture fall into one or another of the classes, or are just left outside of the classifications, and in which settings we can impose a prototypical shape of a Stripe on a sideboard, must be determined by a large number of factors. Among such factors, cultural ones (e.g., what used to be the traditional shape of a cupboard in this culture?) are unlikely to be the least important. We propose a generalization according to which there are no specialized dimensional terms for sideboards, or glasses, or trousers; for various reasons such objects can either be described with some more general terms, or they can stay beyond the scope of our semantic map (in the latter case general size terms, like 'big' or 'small', may be applicable to them). Hence it appears that the typological issues raised by Bierwisch and Lang are undoubtedly important but are in a certain sense of a secondary nature.

6.3   Anna Wierzbicka: Dimensional terms
        in the natural semantic metalanguage

Another extensive study of the semantics of dimensional terms was conducted by
Anna Wierzbicka as a part of a bigger project dedicated to spatial expressions in
general (Wierzbicka 2006). The general purpose of Wierzbicka's work is to give
explications of words of natural language in the Natural Semantic Metalanguage
(NSM). This is a formalized language that consists of semantic primes, which are
claimed to be the universal building blocks of all lexical and grammatical meanings.
Wierzbicka's analysis is based on highlighting minute details and subtle differences
which most semanticists tend to smooth over. For instance, the English adjective
wide can have, according to Wierzbicka, as many as four different senses (each
having its own semantic representation in the NSM):

–   a $wide_1$ chair (seat, bed, sofa, mattress)
–   a $wide_2$ passage (track, corridor, path)
–   a $wide_3$ belt (ribbon)
–   a $wide_4$ street (river)

Wierzbicka claims that another English adjective, *broad* (more specifically, $broad_1$),
is compatible with two types of objects which can hardly be described as *wide*.
One of them are parts of the body which have no definite "edges", and the other
are open spaces like *fields*, *lakes* and *meadows*, i.e. exactly what we call Surfaces.
For Wierzbicka, *a wide field*, at least in the non-metaphorical sense, is impossible.
However, as English NPs like *a wide field of wheat* and *a wide blue lake* are per-
fectly grammatical, we may add a fifth sense of *wide* with the same NSM semantic
representation as $broad_2$.

   As Wierzbicka herself admits, this quinary distinction in senses arises from
pure speculation; it is not corroborated by, say, different linguistic behaviours of
different English *wide*'s. It is then especially interesting that the different behav-
iors fairly neatly correspond to the frames singled out by our typological analysis.
Wierzbicka's five meanings and the four frames on our semantic map (Figure 14)
that are covered with English *wide* correlate in a rather simple way. $Wide_3$ corre-
sponds to the frame we have called WIDE STRIPES, $broad_2$ stands for our BROAD
SURFACES, and $wide_2$ is in substance the same as our WIDE TUBES & HOLES. Objects
which can be $wide_4$ *(streets* and *rivers)* are what we call WIDE ROADS, and the $wide_1$
objects are for us but "degenerate" pieces of a Stripe.

   There are two main discrepancies between our and Wierzbicka's analyses. One
is the borderline between Roads and Tubes in our classification, versus the differ-
ence between Wierzbicka's $wide_2$ and $wide_{1,3,4}$ objects. For Wierzbicka, the crucial

difference between those classes is bodycentricity. For instance, the NSM expli-
cation for wide$_2$ includes the following definition which crucially involves body
placement (Wierzbicka 2006: 156).

> "people can think about things of this kind like this:
>
> "it is a kind of place.
> it has two edges$_{[M]}$
> when a person is moving in this place
> like people move in places of this kind
> one of these two edges$_{[M]}$ is on the side of this person's body
>   where one of the person's arms$_{[M]}$ is
> the other edge$_{[M]}$ is on the side of the body
>   where the other arm$_{[M]}$ is"

Such a passage is absent from the explication of *wide$_4$* (as the edges of a street are
usually far from the arms of the human going along it); thus, *paths* are classified
as different from *streets* (but the same as *corridors*). However, the only language in
our sample that distinguishes WIDE ROADS from WIDE TUBES & HOLES is Nenets,
and for Nenets, it is definitely the form of the objects that distinguishes between
the two WIDE-terms, *tiya* and *pik*. As the examples in (22) and (23) suggest, *tiya*
measures the distance between the two edges of flattish things, and *pik* measures
the space between two sides of hollow things.

(22) TUNDRA NENETS
    a.   *tiya (*pik) latə*
        narrow   board
    b.   *tiya (*pik) sexeri*
        narrow   way
    c.   *tiya (*pik) sarp'ako*
        narrow   path
    d.   *tiya (*pik) ŋuq*
        narrow   trail

(23) a.   *pik (*tiya) xanh m'u*
        narrow   inner.part.of.sledge
    b.   *pik (*tiya) n'o s'i*
        narrow   tent.doorway

However, we are not going to predict that objects like paths and objects like streets
will never be categorized as being dimensionally different in the world's languages.
In a way, a *path*, which often runs between rocks or in the middle of a thick shrub
(entailing higher risk of grazing one's arms or shoulders) is more of a Hole than is a

street. Thus if, in compliance with Wierzbicka's ideas, there is a language where paths behave like Holes and not like Roads, they would still fit into our semantic map.

The other point where our analysis differs from Wierzbica's is whether *wide seats* constitute a special type of context which is cross-linguistically relevant. We would rather speculate that it is peripheral, as the two languages in our sample that lexically distinguish between "bodycentric" and "geometrical" 'wide', Kazakh and Avar, allow both of the relevant lexemes to describe beds and chairs. We claim that this can be attributed to the multiple categorization possibility of these objects. They can be perceived both as "degenerate" pieces of Stripe due to their form, and as elongated places where humans can be placed (due to their function); this is definitely a problem for further research.

6.4   Questions remaining

In this paper, we have discussed the principles of lexicalization of dimensions in some of the languages of Eurasia. A useful tool to formalize these principles is a semantic map, which is a hypothesis about the organisation of the conceptual space, predicting the possible and the impossible patterns of colexification. The number of frames in the field of dimension is small, and the basic semantic map presented in Figure 13 is organised in a rather simple way. The frames are merely ordered in a line, whereas, for example, the semantic map of 'even' (Chapter 6, Figure 1) has a dense network of connections between the frames. Naturally, this implies severer restrictions predicted by our map. However, even this turns out to be not enough for the system to be restrictive enough. We have suggested some additional generalizations which allow restricting the possible lexicalization patterns even further. Thus, the tendencies we have discovered are strong in comparison with most other lexical fields.

Still, we by no means claim that we have exhausted all the typological possibilities of how dimensions may be lexicalized; there may well remain systems that are yet unaccounted. For example, a language may lack several or all specialized dimensional terms altogether. Our sample contains only one language from outside of Eurasia, namely Ch'orti'[7] belonging to the Ch'olan-Tzeltalan group of the Mayan family, which is famous for its elaborated systems of spatial expressions and posture verbs (see Levinson 1997, 2006). The dimensional system in Ch'orti' seems to be radically different from all the other systems in our sample, but in an inverted way: the language almost entirely lacks dimensional terms. The majority of contexts in our questionnaire could only be filled with the general size terms *nuxi' / nixi'* 'big' and *chuchu / chichi* 'small'. In some contexts the diminutive clitic *yar* was

---

7.   We thank Igor Vinogradov for generously providing the Ch'orti data.

also used, but in most contexts there was no specialized word that would imply a particular shape of the object and describe its specific dimension. For example, *nixi' te'* means 'big tree', which implies that the tree is perhaps both tall and thick. Only two words seem to be more or less equivalent to the Spanish dimensional terms. One is *nojta'* 'long'; another is *chilin* 'thin (of Poles)'; both of them are used to describe Poles (ropes, sticks etc.), and *chilin* can be used to describe skinny humans as well. Including more non-Eurasian languages in our sample will allow us to learn whether this "meagre" type of dimensional systems is widespread in the world's languages.

Acknowledgements

References

Abaev, V. 1995. *Izbrannye trudy. Tom II. Obščee i sravniteľnoe jazykoznanie* [Selected Works. Volume II. General and Comparative Linguistics]. Vladikavkaz: Ir.

Bierwisch, M. & Lang, E. 1989. *Dimensional Terms: Grammatical Structure and Conceptual Interpretation*. Berlin, Heidelberg: Springer-Verlag

Brown, D., Chumakina, M. & Corbett, G. G. (eds). 2013. *Canonical Morphology and Syntax*. Oxford: Oxford University Press.

Chirikba, V. 2008. The problem of the Caucasian Sprachbund. In *From Linguistic Areas to Areal Linguistics*, P. Muysken (ed), 25–94. Amsterdam, Philadlphia: John Benjamins. https://doi.org/10.1075/slcs.90.02chi

Dirven, R. & Taylor, J. 1988. The conceptualisation of vertical space in English: the case of ''tall''. In *Topics in Cognitive Linguistics*, B. Rudzka-Ostyn (ed), 207–229. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.50.15dir

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language* 67(3): 547–619. https://doi.org/10.1353/lan.1991.0021

Fillmore, Ch. J. 1978. On the organization of semantic information in the lexicon. In *Papers from the Parasession on the Lexicon*, 148–173. Chicago: Chicago Linguistic Society.

Fillmore, Ch. J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, 111–137. Seoul: Hanshin.

Greenberg, J. H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, J. H. Greenberg (ed), 73–113. Cambridge, Massachusets & London, England: MIT Press.

Haspelmath, M. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In *New Psychology of Language, vol. 2*, M. Tomasello (ed), 211–242. Mahwah, NJ: Lawrence Erlbaum.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(4): 663–687. https://doi.org/10.1353/lan.2010.0021

Jackendoff, R. & Landau, B. 1993. "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16: 217–265. https://doi.org/10.1017/S0140525X00029733

Jones, S., Murphy, M. L., Paradis, C. & Willners, C. 2012. *Antonyms in English: Construals, Constructions and Canonicity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139032384

Klimov, G. (ed). 1978. *Strukturnye obščnosti kavkazskix jazykov* [The Structural Affinities Between the Caucasian Languages]. Moskva: Nauka.

Koptjevskaja-Tamm, M. & Miestamo, M. 2015. Antonyms and wordlevel negation. Paper presented at Diversity Linguistics: Retrospect and Prospect, Max Planck Institute for Evolutionary anthropology, Leipzig, 2 May 2015.

Kukhto, A., Kozlov, A. & Privizentseva, M. 2016. O real'nosti semantičeskogo pol'a: leksiko-tipologičeskij podxod [On the reality of a semantic field: lexico-typological approach]. *Acta Linguistica Petropolitana* XII (1): 522–533.

Lakoff, G. & Johnson, M. 1999. *Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lang, E. 2001. Spatial dimension terms. In *Language Typology and Language Universals. An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1251–1275. Berlin: Mouton de Gruyter.

Levinson, S. C. 1997. Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology* 7(1): 98–131. https://doi.org/10.1525/jlin.1997.7.1.98

Levinson, S. C. 2006. The language of space in Yélî Dnye. In *Grammars of Space. Explorations in Cognitive Diversity*, S. C. Levinson, & D. Wilkins (eds), 157–203. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486753.006

Murphy, M. L. 2003. *Semantic Relations and the Lexicon: Antonymy, Synonymy, and other Paradigms*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486494

Paradis, C., Willners, C. & Jones, S. 2009. Good and bad opposites: using textual and psycholinguistic techniques to measure antonym canonicity. *The Mental Lexicon* 4(3): 380–429. https://doi.org/10.1075/ml.4.3.04par

Pulvermüller, F. 2005. Brain Mechanisms Linking Language and Action. *Nature Reviews Neuroscience* 6: 576–582. https://doi.org/10.1038/nrn1706

Rakhilina, E. V. 2000. *Kognitivnyj analiz predmentyx imen: semantika i sočetaemost'* [Cognitive Analysis of Concrete Nouns: Semantics and Combinability]. Moscow: Russkie slovari.

Redei, K. 1986. *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai. Kiadó.

Talmy, L. 1983. How language structures space. In *Spatial Orientation: Theory, Research and Application*, H. L. Pick & L. P. Acredolo (eds), 225–282. New York: Plenum Press. https://doi.org/10.1007/978-1-4615-9325-6_11

Varela, F. J., Thompson, E. & Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6730.001.0001

Wierzbicka, A. 2006. Shape in grammar revisited. *Studies in Language* 30(1): 115–177. https://doi.org/10.1075/sl.30.1.05wie

Zaliznyak, A. A. 1985. *Ot praslavyanskoy akcentuacii k russkoy*. [From the Proto-Slavic Accent System towards the Russian one]. Moscow: Nauka.

The domain of surface texture

Egor Kashkin and Olga Vinogradova

Vinogradov Russian Language Institute of the Russian Academy of Sciences /
HSE University, Moscow

The paper deals with the typology of surface texture expressions, such as
slippery *road,* ***smooth*** *wooden board,* ***rough*** *hands,* ***coarse*** *or* ***rough*** *fabric*. We
discuss both their literal uses and metaphors formed with them, such as ***slip-***
pery *person,* ***smooth*** *speech,* ***rugged*** *captain*. Our language sample includes 10
Uralic languages (Finnish, Estonian, Meadow Mari, Erzya, Moksha, Udmurt,
Komi-Zyrjan, Hungarian, Khanty, Nenets), as well as 5 languages from other
families (Russian, English, Spanish, Chinese, and Korean). The categorisation
includes primarily a division into visually perceived surfaces and surfaces per-
ceived through physical contact. We discuss in what ways the antonymic areas
under observation are asymmetrical in their semantics and combinability. One
more focus is on evaluating variation in the texture lexicon in genetically related
languages in comparison with its variation across a broader sample of languages.

Keywords: lexical typology, intragenetic typology, corpus research, metaphoric
shift, Uralic languages

1.   Introduction

The work presents an attempt to categorise one part of the domain of sensory
lexicon, namely, attributes for surface texture description. Having ten Uralic lan-
guages (Finnish, Estonian, Meadow Mari, Erzya, Moksha, Udmurt, Komi-Zyrjan,
Hungarian, Khanty, Nenets) and five languages from other families (Russian,
English, Spanish, Mandarin Chinese, and Korean) in our language sample might
look like a disproportionate selection, but the sample has not been compiled ran-
domly. Our aim was, first, to find out the degree of similarity demonstrated by the
same semantic domain in a group of closely related languages, and, second, to check
the typological relevance of our results on languages outside this family. The Uralic

data have been mostly gathered in fieldwork,[1] but materials from dictionaries and corpora were also included where available.[2] The Russian material (Russian being the native language of the authors) comes primarily from the Russian National Corpus (RNC) and from dictionaries. Our English data were collected by getting responses of native speakers to our typological questionnaires, as well as by consulting the Corpus of Contemporary American English (COCA),[3] the British National Corpus (BNC),[4] and a range of online and paper dictionaries. As regards the other languages in our sample, their data have been taken from the existing descriptions created within our project, namely: (Spesivceva 2012) for Spanish, (Ivanova 2011) for Korean, and (Kholkina 2014) for Mandarin Chinese.

The data on some particular languages were published in our previous papers, see Kashkin (2011a, 2011b, 2012, 2013a: 112–280, 2013b), Vinogradova (2013) and Koshkareva et al. (2017: 233–242). In this paper we aim to discuss cross-linguistic generalizations of our study, providing language examples for them, but not elaborating on the whole structure of the surface texture domain in each particular language.

The description of the research results falls into four sections. After the introduction in Section 1, we discuss the literal meanings of surface texture expressions (covering both roughness and its absence) in Section 2. Section 3 is devoted to the metaphorical uses derived in the domain in question. Finally, the possible conclusions and implications are considered in Section 4.

2.   Literal uses

2.1   Absence of roughness

2.1.1   *The way a surface is perceived*
The basic semantic distinction within the domain of surfaces deals with **the way a surface is perceived**.[5] For some surfaces their 'smoothness' or 'roughness' is evaluated **visually**, like in the case of a field or a ceiling (we will use the label LEVEL for

---

1.   If an example receives no explicit reference, it means that it has been recorded from a native speaker.

2.   Certain Uralic dialects different from the standard language variety are referred to when discussing the relevant examples.

3.   https://www.english-corpora.org/coca/, last accessed October 1, 2019.

4.   https://www.english-corpora.org/bnc/, last accessed October 1, 2019.

5.   Here and further on we discuss linguistic understanding of perception. For the biological aspects of this issue, see (Lederman & Klatzky 2009) and references therein.

the group of frames[6] that includes fields and ceilings). There are, however, many surfaces usually perceived **by touch**, cf. a wooden board when one examines how well it is polished, or one's cheeks while shaving. The frames of tactile perception are further subdivided into two types: the first one includes **smooth** surfaces like those mentioned above, and the second one embraces **slippery** surfaces (e.g. a slippery road or a slippery ball). Slippery surfaces lack roughness to such an extent that it becomes difficult to keep one's balance on them or to hold them in one's hands. Correspondingly, the labels we use are SMOOTH and SLIPPERY.

The languages of our sample demonstrate three strategies of categorising SLIPPERY, SMOOTH, and LEVEL surfaces.[7]

First, each of these types may be referred to by a special lexeme (or a special set of lexemes), cf. Examples (1)–(3) from Russian. SLIPPERY surfaces are described with an adjective *skol'zkij* (1). SMOOTH surfaces (perceived by touch) require another adjective *gladkij* (2). Finally, LEVEL surfaces are referred to as *rovnyj* (3).

RUSSIAN

(1) *Ja podnja-l-sja   i,   ostorožno stupaj-a   po **skol'zk-omu***
    I lift-PST.M.SG-REFL and carefully tread-CVB.PRS on slippery-DAT.M.SG
    parket-u,   otpravi-l-sja   v kuxnj-u.
    parquet-DAT.SG send-PST.M.SG-REFL to kitchen-ACC.SG
    'I stood up and went to the kitchen, carefully treading on a slippery parquet floor'
    [RNC]

RUSSIAN

(2) *Poverxnost'   kartin-y   dolžn-a   by-l-a   by-t'*
    surface.NOM.SG picture-GEN.SG must-F.SG be-PST-F.SG be-INF
    ***gladk-oj**,   kak polirovann-aja   kost'.*
    smooth-INS.F.SG as polished- NOM.F.SG ivory.NOM.SG
    'The surface of the picture was required to be as smooth as polished ivory'
    [RNC]

RUSSIAN

(3) *Rjadom s   dorog-oj   polno **rovn-yx**   ploščadok, jest'*
    close   with road-INS.SG plenty level-GEN.PL area.GEN.PL be.PRS
    rodnik-i   i   ruč'j-i.
    spring-NOM.PL and brook-NOM.PL
    'There are plenty of level areas near the road, as well as springs and brooks'
    [RNC]

---

6. We use the term "frame" for a basic situation described by a lexeme, see Rakhilina and Reznikova (this volume). One can also speak about groups of frames in the sense that frames in a group are semantically close to each other and are often co-lexified.

7. It should be emphasized that SLIPPERY, SMOOTH, and LEVEL are used here as metalinguistic labels referring to the basic subzones of surfaces without roughness, and they do not imply all the polysemy of the corresponding English lexemes.

Second, languages sometimes use a separate lexeme for SLIPPERY surfaces in op-position to just one lexeme in common for both SMOOTH and LEVEL surfaces. This strategy can be illustrated in Erzya Examples (4)–(6): the adjective *nolaža* stands for SLIPPERY (4), while the adjective *valan'a* describes both SMOOTH (5) and LEVEL (6) surfaces.

ERZYA

(4) *vasn'a son (nad'a) pel'-s'   synst   (kal-tne-n')   tokše-ms:*
     at.first she Nadya be.afraid-PST.3SG they.GEN fish-PL.DEF-GEN touch-INF
     nolaža-t   di jezmoldy-t'
     slippery-PL and move-NPST.3PL
     'At first she (Nadya) was afraid of touching them (small fishes): they were
     slippery and were moving'   (Buzakova 1982: 88).

ERZYA

(5) ***valan'a** ked'-se   son panar vi-kšn'-i*
     smooth hand-LOC she shirt sew-IPFV-NPST.3SG
     'She is sewing a shirt with her smooth hands'

ERZYA

(6) ***valan'a** paks'a-nt'   langa s'ed'e vadr'a kizna   t'ikše-n'*
     level   field-DEF.GEN on   more good   summer grass-GEN
     l'ed'e-ma-s'
     mow-NMN-DEF.NOM
     'In summer it is better to mow grass on a level field'

In the third strategy, SLIPPERY and SMOOTH surfaces can be described with one and the same lexeme, while LEVEL surfaces are referred to with another lexeme. This strategy is less frequent than the other two, but at the same time it provides a typo-logical reason for including SLIPPERY surfaces into the same larger domain (absence of roughness) as SMOOTH ones. An illustration is represented in Examples (7)–(9) from Khanty (Tegi dialect): an adjective *woł'ək* (with an allomorph *woł'k*) means 'slippery' (7), and 'smooth' (8), whereas an adjective *pajłi* means 'level' (9). Outside the Uralic family, this strategy is also attested in Chinese (Kholkina 2014: 170–171).

KHANTY (TEGI DIALECT)

(7) *at   măr juoš pɔt-s-a,   juoš **woł'k-a**   ji-s*
     night at   road freeze-PST-PASS road slippery-DAT become-PST
     'The road has frozen at night, it has become slippery'

KHANTY (TEGI DIALECT)

(8) *tăm sɔxəł   **woł'k-a**   wer-e!*
     this wooden.board smooth-DAT make-IMP.O.SG
     'Make this wooden board smooth' (e.g., a father tells this to his son while
     teaching him to polish wooden boards).

Khanty (Tegi dialect)

(9) *tăm torən xăr-em*   ***pajłi****, śit-em*   *muw-əŋ.*
      this hay   place-POSS1SG level that-POSS1SG ground-ATTR
      'This meadowland of mine is level, and that one is hummocky'

These are the basic classes of surfaces "without roughness" and the strategies languages use in subdividing this semantic zone. These semantic oppositions within the domain concerned, however, are not limited to the distinctions among these three basic classes. Each of these classes includes some rather heterogeneous frames, which is apparent in the data investigated. In the next subsections we will look at more semantic features, starting with the subdomain of SLIPPERY surfaces.

2.1.2   *Slippery surfaces*

The crucial distinction among different SLIPPERY surfaces deals with their **topological class**, which primarily implies the type of contact between a surface and a human being. On the one hand, one may evaluate a surface as SLIPPERY when moving on it and finding it difficult to keep one's balance (e.g., a road in winter, a wet floor, or stone stairs). On the other hand, one may try to keep an object with a slippery surface in one's hands and experience problems trying to hold it (e.g., fish, a bar of soap, or the handle of a spade). There is also a frame that is in a way similar to both topological classes of slippery surfaces, and a typical representative is the lexeme denoting the sole of shoes (and metonymically shoes themselves). Shoe soles may slip as one is moving (which makes them similar to the road), and at the same time the sole is small and its topological properties therefore become close to those of fish or small stones.

Quite a few languages possess a lexeme dominant over all the frames of SLIPPERY surfaces, cf. English *slippery*, Russian *skol'zkij*, Erzya *nolaža*, and Estonian *libe*. Some languages (e.g., Komi, Udmurt, Spanish), however, treat bearing surfaces separately from the surfaces of objects slipping out of hands, cf. the Izhma Komi Examples (10) and (11). In (10) a slippery floor is described with an adjective *vol'k* 'slippery', while in (11) this adjective is out of place for describing a piece of soap, and what must be used instead is a verb *vol'sjoony* 'to slip out'.

Komi (Izhma dialect)

(10) *pos-se*   *mys'k-ema-s', i   l'ok-a   kos't-ema-s', i*
       floor-ACC.POSS3SG wash-PST2-PL and bad-ADV dry-PST2-PL and
       posk-ys   ***vol'k****,   verm-an   us'-ny.*
       floor.OBL-POSS3SG slippery may-NPST.2SG fall-INF
       'The floor has been mopped but badly dried, the floor is slippery, you may fall
       down'

Komi (Izhma dialect)

(11) *majteg torj-ys*   **vol'sjal-e**   / ***vol'k.**
      soap   piece.OBL-POSS3SG slip.out-PRS.3SG   slippery
      'The piece of soap is slipping out'

The frame of shoe soles may behave differently in such systems. In Izhma Komi, it can be characterized either with the adjective *vol'k,* or with the verb *vol'sjoony*. However, in the closely related Udmurt language, it is only the verb *gyldz'yny* 'to slip out' (about fish, soap, etc.) that is applied here, but not the adjective *gylyt* used for bearing surfaces. In Spanish, the lexeme *resbaladizo* is common for both bearing surfaces and shoe soles, whereas a lexeme *escurridizo* is specific for objects slipping out of hands (Spesivceva 2012: 39–42).

Apart from the differences determined by its topological properties, a surface can be slippery for different reasons. For the sole of a shoe this is usually an inherent property, but at the same time many surfaces can become slippery because there is some **extraneous substance** on them, such as ice on the road, slime on the skin of a snake or on the scale of a fish. This factor becomes significant for some languages. In COCA, the 101 most frequent occurrences of the English adjective *slick*[8] are in 64 combinations with nouns, and 43 of them are nouns for objects that have become slippery because of something that happened to them, such as a road with thawing snow or a path covered with fallen leaves. Here are some of those occurrences: *Ahead, the trail rose steeply, **slick** stones and black mud churned by booted feet and shod hooves; … with his bare torso **slick** with water and mud from the damp,…; … the windows overlooking the quad **slick** with condensation, the children overlooking each other **slick** with sweat,…; I noticed the snail trail of **slick** tears across his cheeks.; … a dead fruit fly floating belly-up in the rainbow-colored grease **slick** spreading on the surface of his coffee …; … He nearly slipped as his foot hit something **slick**. He looked down briefly to see the puddle of crimson he'd skidded through.*

Similarly, Hungarian *síkos* is used only for slippery surfaces covered with extraneous substances. Its prototypical contexts are an icy road, the skin of a snake, and the scale of a fish (on the contrary, the adjective *csúszós* is dominant over all the subdomains of SLIPPERY surfaces in Hungarian).[9]

---

8. https://www.english-corpora.org/coca/, search for *slick*, last accessed October 13, 2019.

9. It should be noted here that we do not analyze adjectives like *slimy* (often derived directly from a noun referring to the extraneous substance), as they do not necessarily imply that an object is slippery, often focusing only on the existence of *slime* on its surface, as in the example from COCA: … *I tried to climb out of the hole full of silt and rotting vegetation and maybe **slimy** creatures.*

A special strategy is represented in some languages which have a separate lexeme for a slippery **bearing surface covered with ice** (e.g., a road after it has frozen at night), and this lexeme is not a direct derivation from the noun 'ice'. In particular, this is the case of Tundra Nenets, where a lexeme *saløt°q* means 'slippery with ice' (about a bearing surface), while the dominant lexeme for all slippery surfaces is a verb *nøsadørcy°* 'to slip'.

2.1.3   *Smooth surfaces*

Smooth surface semantics cannot always be reduced to the type of perception, as lexemes sometimes include more fine-grained distinctions. One important type of smooth surface includes the **additional visual feature of shining**. A striking example of lexicalizing this feature is provided by an English adjective *sleek* meaning 'smooth and shining' (cf. the neutral term *smooth*). Examples can be seen in photographs of *sleek* hair on the internet – the hair is typically very smooth and shining, reflecting light in a special way, often as a result of special care or styling, and in most cases such hair is of quite a "straightforward" shade giving the strongest luster, mostly raven-black or golden. Here are a few other examples of *sleek* used in the same meaning: *He wore an old **sleek** crisp flaxen wig which …was made of hair but which looked far more as though it was spun from filaments of silk or glass. (Charles Dickens, A Tale of Two Cities); …a few styling tricks…the result: **sleek** face; …the actress's **sleek** face flaming with tints of arousal…; Light glints off the glass tiles, drawing the eye to their **sleek** surfaces.*

A bit more complex case is observed in Spanish, where the adjective *terso* describes only sleek surfaces (Spesivceva 2012: 35–37), at the same time imposing restrictions on the semantic class of the object: it is compatible only with the names of body parts (e.g., hands or cheeks), but not artefacts (e.g., a polished wooden board). Within the Uralic family, the parameter of 'shining' is relevant for Meadow Mari: an adjective *jaklaka* refers to all the slippery surfaces, including sleek surfaces, whereas smooth surfaces which do not reflect light are described as *jyvyža*.

As was demonstrated with the Spanish adjective *terso*, some lexemes may describe only body parts, which therefore constitute a special type of surface. In some languages, this determines a special strategy of categorising **body parts**. The focus in their description may be placed on some other related features, while lexemes of surface texture cannot be applied to them. This strategy has been attested in Khanty (Muzhi dialect) and in Moksha (Central dialect). In Khanty an adjective *pajłi* 'level, smooth' may describe smooth artefacts (e.g., wooden boards), but not a person's skin, face, or hair. The qualities of these body parts are categorised within other semantic domains, e.g., *mułijəł* (shine-NPST.3SG) 'shines' or *xɔrasey* 'beautiful' about a man's clean-shaven face, or about a smooth face of a pretty girl; *łepət* 'soft' about well-groomed skin.

2.1.4   LEVEL *surfaces: Land areas vs. artefacts*

A detailed language categorisation of perception types emerges not only for SLIP-PERY and SMOOTH surfaces, but also for those in the domain of LEVEL. The first stage of this categorisation deals with the opposition between level **land areas** and **artefacts**. Thus, Hungarian *sík* 'level, flat' is appropriate only for land areas, for example, *sík rét* 'level meadow', *sík mező* 'level field', *sík vidék* 'level landscape'. Izhma Komi gives a more complex case, where an adjective *mol'yd* describes all the smooth surfaces perceived by touch, and also level land areas, but not level artefacts (e.g., it is not applied to a floor, a wall, a fence) if a speaker means their visual perception. The latter class requires an adjective *rəvnej* (borrowed from Russian *rovnyj* 'level') or an adjective *ves'kyd*, which means 'straight' and which may also cover some frames of artefacts.

The class of land areas shows more fine-grained distinctions which cannot be reduced to a mere type of perception. Thus, a special frame in this subdomain is a **level road**. On the one hand, it is a land area and is often evaluated visually; on the other hand, a road has a specific function of being a surface on which people move. This makes it subject to a certain degree of tactile evaluation, which affects the choice of texture expressions for it. Languages differentiating between the lexemes for 'level' and 'smooth' tend to use the latter for a road, especially if it is necessary to emphasize the good quality of a road (while an adjective 'level' still remains possible). For example, the Russian National Corpus provides 100 entries of *rovnaja doroga* 'a level road' and 63 entries of *gladkaja doroga* 'a smooth road' asserting a greater degree of 'smoothness' for those who use it. For other kinds of level surfaces, the occurrence of the 'tactile' adjective is considerably less frequent. Thus, the ratio of co-occurrences of *rovnyj* 'level' as opposed to *gladkij* 'smooth' is 84 vs. 14 with *pole* 'field' (the latter examples all come from fiction between the 18th and 20th centuries), 46 vs. 9 with *step'* 'steppe' (the last example for *gladkij* dates back to 1937), and 14 vs. 2 with *lug* 'meadow' (both of the entries of *gladkij* with *lug* date back to the 19th century).

One more important frame within land areas is landscape **without mountains** or heights. Some languages encode it with a lexeme originally coming from the semantic domain of shape and denoting flat objects, e.g. Russian *ploskij* or Estonian *lame*. However, this polysemy pattern is not universal. In many languages, a lexeme describing a flat shape cannot refer to any surface landscape properties at all (like Khanty *lopsəx* or Moksha *lapš*). On the other hand, some lexemes traced back to the domain of shape are applicable not only to landscapes, but also to other types of surfaces. This is the case of English *flat*, which is applicable to various kinds of surfaces defined by one of our consultants as having no "significant elevation"; this is also seen in the following examples from COCA: *a perfectly **flat** floor for a dining*

*table and six chair*s; *Stand the chair on a* **flat** *table and check that all legs rest evenly*; *The top of the mesa was a* **flat** *deck of stone.*

In the subdomain of "land" areas, there is the frame of smooth **water surface** (when there is no wind). Some languages use lexemes here from the surface domain, e.g. English *smooth*, Russian *gladkij*, or Nenets *salmuy*. It is often the case, however, that water cannot be described with surface expressions. Instead, some languages focus on other related features in its categorisation, see Example (12) from Tegi Khanty, or an Udmurt expression *šypyt pukys'* (lit.: quietly sitting) used about smooth water surface.

Khanty (Tegi dialect)

(12) *tewən, jiŋk ńur   rɔm-ije.*
    quiet   water entirely calm-DIM
    'Be quiet, the water is entirely calm'

Artefacts with a level surface may be differentiated on the basis of their **vertical** vs. **horizontal orientation**. A good example is provided by Tegi Khanty, where an adjective *pajłi* 'level' is appropriate only for horizontal surfaces (e.g., a floor), while vertical surfaces (e.g., a wall or a fence) tend to be described as *tuŋ* 'straight'.

The typological data on how the absence of irregularities may be categorised across languages are summarized on the semantic map in Figure 1. The map has been created manually following the theoretical approach of Haspelmath (2003).
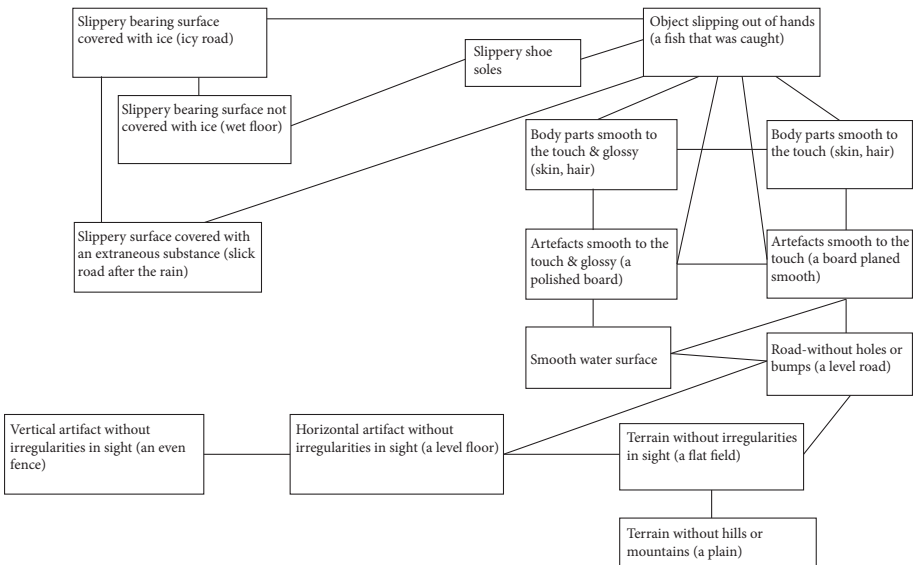
Figure 1. Absence of irregularities domain: basic semantic map

2.2   Roughness

As compared to the zone of smoothness, the frames of roughness have an additional slot, which is roughness itself with its own properties: size, regularity, rigidity. As we will see, this becomes highly important for language categorisation of ROUGH surfaces. The zone of roughness includes quite a few lexemes with narrower semantics referring to various specific types of roughness, e.g. English *rippled, pimply, scarred, knotty, bumpy, cracked, bristly*. We will not discuss such lexemes in this article, but we will focus on classes of frames typically categorised by lexemes with long collocation lists.

According to our cross-linguistic data, there are two important classes of ROUGH surfaces. The first one embraces wrinkled surfaces, like the face of an elderly person or the skin of an old apple, cf. Udmurt *kisyrijo*, Khanty *mormǝŋ*, Izhma Komi *kǝrs'ema* as examples of lexemes specific for this class. Sometimes wrinkled surfaces are subdivided into two-dimensional and three-dimensional objects, cf. Estonian *kortsus* 'wrinkled' (for a flat object such as skin, face, forehead, leaf) vs. *krimpsus* 'wrinkled' (for a 3-dimensional object such as an apple, potato). The second important class of ROUGH surfaces comprises surfaces with regularly rigid roughness perceived by touch such as a cat's tongue, frost-bitten hands, or a badly polished piece of wood (e.g. Russian *šeršavyj*, Udmurt *šakyres*, Erzya *kaz'amo*, and Estonian *kare* correspond to this kind of surfaces). It is the latter class which demonstrates the most prominent cross-linguistic variation along with having many of the prototypes mentioned above that are common for the languages in our sample. We will therefore concentrate on its typology.

ROUGH surfaces may vary in the **size of roughness**: on the one hand, small evenly-distributed granules of roughness on a cat's tongue or on frostbitten hands, or, on the other hand, larger patch-like roughness like that of the tree bark or a scab. Within our set of Uralic languages, this opposition is significant for Finnish *karhea* (small roughness) vs. *karkea* (large roughness), and for Estonian *kare* (small roughness) vs. *krobeline* (large roughness). This distinction is illustrated by the Estonian Examples (13)–(14), with the latter example given by a native speaker in reply to the question if there is any situation when the surface of paper or another similar object may be called *krobeline*.

ESTONIAN

(13) *See   paber   on   natuke **kare**   / *****krobeline***.
      this paper be.PRS.3 a.bit   rough   coarse
      'This paper is a bit rough' (about paper in an old notebook)

ESTONIAN

(14) *Krokodillinaha*   *imitatsiooni-ga vihikukaane-d*   *on*
      crocodile+skin:GEN imitation-COM notebook+cover-PL be.PRS.3
      kergelt **krobelise-d.**
      a.bit   coarse-PL
      'Notebook covers imitating crocodile skin are a bit coarse'

Outside the Uralic family, a good example of this opposition is provided by the group of English adjectives *rough, coarse*, and *rugged*. The last lexeme obviously refers to large roughness, often located on landscapes or other visually perceived objects, e.g. *the lava's* **rugged** *surface provided livestock with a good place to break their legs* [COCA]*; Phobos has … a more* **rugged** *surface: its most striking features are a large impact crater and a series of grooves* [COCA]. *Coarse*, in its turn, tends to denote more significant irregularities than *rough*. Often it emphasizes a greater degree of roughness, like in the example *Maine is a vast and empty state, and that is precisely what Steve loves most about it. I understand the appeal on a theoretical level, but in actuality I find these woods* **coarse** *and lonely.* [COCA]. Another illustrative example of the meaning is how these two adjectives are used to describe whetstones or sandpaper. As they are by definition *rough*, they are rarely characterized with this adjective, while the use of *coarse* is common with them if the speaker focuses on the larger size of grains.

The parameter of size sometimes correlates with other properties of a surface, such as size of granules and positive versus negative evaluation. First, if we look at the combinations of *rough* and *coarse* with the same nouns, we notice that the use of *rough* often emphasizes lack of wood surface processing (*rough boards* as in *"… ordered them to cut timber and split it into* **rough** *boards",* or *rough wood planks* as in *"He and I were sitting on the* **rough** *wood slats of the dock…"*), or lack of care (*Comb her hair, and she becomes a young socialite. As it stands, the bold contradiction of* **rough** *hair plus the refined mouth creates a Kate who's not sure whether she's coming or going*). Secondly, *coarse* can highlight the separate structure of the elements contributing to the roughness of a surface or its fibre (*…his chest's* **coarse** *hairs looked golden, his forehead's rugged creases appeared less defined; I took off my loafers and socks and walked out on the hard* **coarse** *surface of the wet sand*). Third, *rough* can communicate the idea of low functionality, though it is important to note that *rough* is still applicable to certain functional and positively evaluated surfaces, cf. *Slightly* **rough** *blankets are also preferred as they have less contact with the paper surface thus reducing blanket contamination* [Google]. This brings us to a fourth observation, that the negative evaluation in **rough** *hair* is rather a frequent pragmatic implication than a necessary assertion, and the primary semantic parameter regulating the use of *rough* is size of roughness. The typological data do not show any lexeme

specifying either positive or negative evaluation of rough surfaces. Overall, the parameters of evaluation and functionality always remain supplementary to any other parameters discussed in this section.

Another parameter of variation within the ROUGH subdomain is the **regularity of roughness**. Thus, the Russian adjective *šeršavyj* refers to roughness which is regularly distributed on a surface and/or is inherent (e.g. on such surfaces as a cat's tongue, asphalt, an emery board), while the adjective *šeroxovatyj* tends to describe irregular roughness (e.g. on a wooden board or tree bark) that is often expected to be treated. In many languages, a lexeme used for regular roughness cannot be applied to a surface with irregular roughness, such as Izhma Komi *sozores'*, Khanty *karəŋ*, or Nenets *nasortøsy*. On the contrary, some languages make no lexical distinction in the regularity of roughness; this is the case of Erzya *kaz'amo*, English *rough*, or Chinese *cucao,* all of which describe roughness irrespective of its regularity.

As mentioned above, a prototype of ROUGH surfaces includes those covered with rigid roughness. There are, however, some soft surfaces covered with separate items which cause what may be called **softer roughness**. Examples include stubble and a woolen blanket: *Still, nervousness itched at him just like the **rough** blanket* [COCA]; *The blankets itched me horribly. My skin felt raw where I had scratched at my arms in my sleep, and when I finally hauled my legs from under the **rough** blankets…* [Google]. Languages differ as to whether such surfaces may be described with a basic lexeme meaning 'rough'. The use of Estonian *kare*, Erzya *kaz'amo* or English *rough* spreads to the frame of soft roughness, while Izhma Komi *sozores'* or Western Khanty *karəŋ* are impossible in these contexts, remaining specific for rigid roughness.

A special class of ROUGH surfaces includes surfaces **affecting an object in contact,** e.g. scratching or pricking it, like bristly cheeks or splintery wood. This is the case of Udmurt *čogyres* which describes surfaces which are simultaneously rough and scratching (15), compared to the neutral term *šakyres* 'rough'. A similar case has been attested in Spanish; see Spesivceva (2012: 55–57) for a detailed discussion of the adjective *rasposo* 'rough and scratching' as opposed to *áspero* 'rough'.

UDMURT
(15) ***čogyres** pul-tɨ*   *basma-en   ortč'yt-ɨ-d   ke, basma-len*
      rough   wooden.board-PROL cloth-INSTR pass-PST-2SG if   cloth-GEN
      sin'ys-jos-yz   kan'žas'k-o-zy   pul   bord-y
      thread-PL-POSS3SG catch.on-FUT-3PL wooden.board near-ILL
      'If you pass a cloth over a rough wooden board, the threads of the cloth will catch on the wooden board'

Most cross-linguistic distinctions in the subdomain of ROUGH discussed so far in this section are connected with the properties of roughness itself. However, some languages maintain lexical differentiation of the objects, namely, the distinction between **natural objects** and **artefacts**. As regards the zone of ROUGH, this can be observed in Mandarin Chinese (Kholkina 2014: 208), where the adjective *máocao* 'rough' is applied to man-made cloth or wooden boards, but cannot refer to objects that exist in wild nature (e.g. tree bark, branches, leaves) and have not been processed by the man.

An important parameter of cross-linguistic variation in the ROUGH subdomain deals with its relation to **coarse fabric.** Languages differ in whether there is a texture lexeme covering this frame. Sometimes there is no such lexeme in the subdomain of ROUGH, cf. Russian *šeršavyj* characterizing a surface vs. *grubyj* describing the stiff structure of any object. In Udmurt, coarse fabric is described with an adjective *č'uryt* literally meaning 'hard, rigid', but not with an adjective *šakyres* 'rough'. Some languages, however, apply the same lexeme to rough surfaces, cf. Shoksha Erzya *kaz'ama* in (16)–(17).

ERZYA (SHOKSHA)

(16) *katka-t'*   *kel'-s'*   **kaz'ama,** *son čama-t'e-st kišt-e*
     cat-DEF.GEN tongue-DEF.NOM rough   it   face-DEF-EL lick-NPST.3SG
     'A cat's tongue is rough, it is licking my face'

ERZYA (SHOKSHA)

(17) *keskaf-t'n'i-n'*   *t'ij^h-sy-z'*   **kaz'ama** *mat'er'ial-sta*
     sack-PL.DEF.GEN make-PRS-3PL.S:PL.O coarse   fabric-EL
     'Sacks are made of coarse fabric'

The semantic map visualizing the domain of ROUGH is shown in Figure 2. Note that the semantic maps of the two antonymic domains (absence of roughness in Figure 1 vs. roughness in Figure 2) are quite different. In particular, lexemes denoting absence of roughness draw subtle distinctions within various types of perception, while lexemes describing roughness are mostly opposed in the properties of roughness itself (size, regularity, impact on a contacted object).
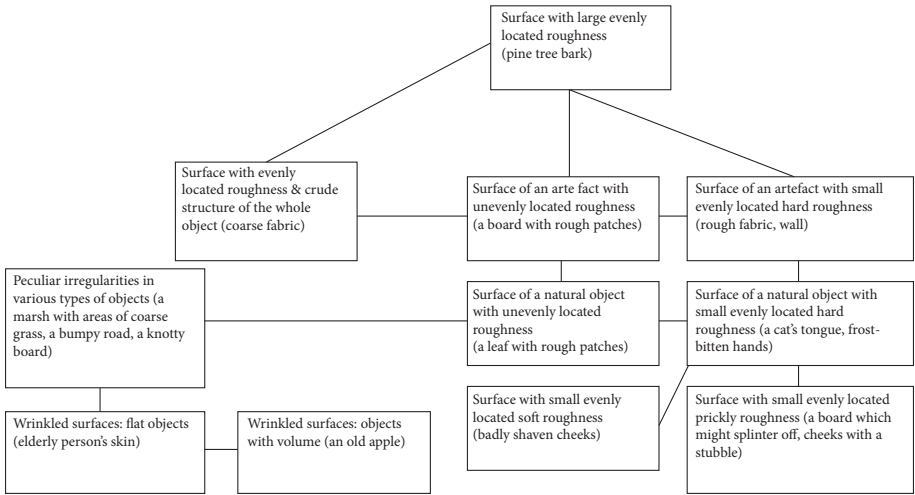
Figure 2. Irregularities domain: Basic semantic map

Figures 3 and 4 illustrate the difference between the two domains as lexicalized in Hungarian.
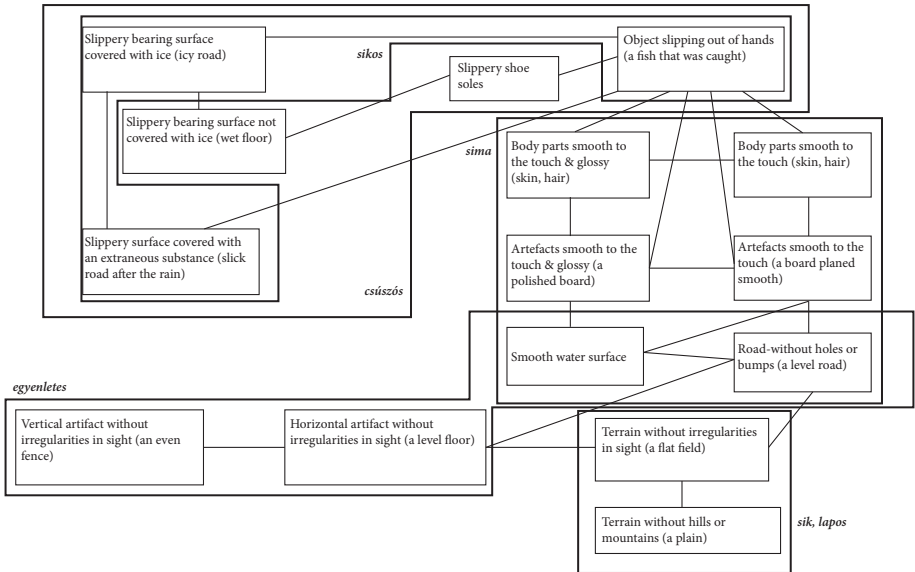
Figure 3. Absence of irregularities domain: Semantic map for Hungarian
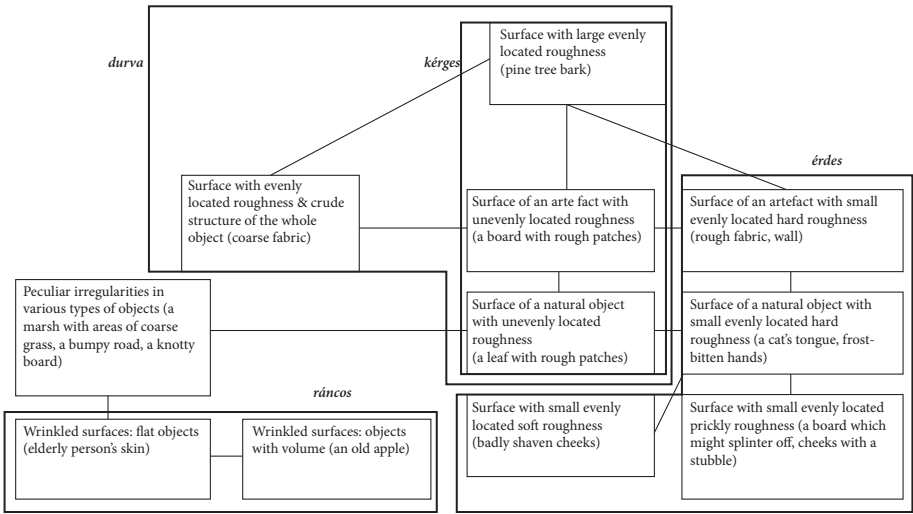
Figure 4. Irregularities domain: Semantic map for Hungarian

3. Metaphoric uses

The analysis of metaphoric uses of the texture expressions will follow the same order of the source semantic zones as they were discussed in Section 2 – SLIPPERY, SMOOTH, LEVEL, and the zone of roughness.

3.1   SLIPPERY

Lexemes describing slippery surfaces often develop metaphors of **unsteadiness or lack of trustworthiness**. This shift is quite transparent, keeping in mind low functionality feature of slippery surfaces, which makes it difficult to use them. Metaphors belonging to this class first refer to unreliable people, like English *a slick person*, *a slick lawyer*, *slick Willie and slippery Hillary* (a reference to Bill and Hillary Clinton; the difference between the two adjectives *slick* and *slippery* in their metaphoric uses is quite subtle; according to the Macmillan dictionary,[10] "a slick person is clever and good at persuading people but probably not honest or sincere," while "someone who is slippery is clever but dishonest, so that you cannot trust them"). Some specific expressions include English *be as slippery as an eel*, Russian *skol'zkij*

---

10. Macmillan English Dictionary for Advanced Learners, International Student edition 2006, p. 1344 and p. 1346.

tip 'a slippery person', Finnish *luikas kuin ankerias* 'as slippery as an eel' (about an unreliable person), and Erzya *nolaža loman'* 'a slippery person'.

Second, metaphors involving 'slippery' may deal with situations, course of life, and fields of somebody's activity which may lead to something unpredictable, e.g. Spanish *situación resbaladiza* 'an unstable (lit.: slippery) situation', Russian *skol'zkaja tema* 'an unpleasantly tricky (lit.: slippery) subject', *skol'zkij vopros* 'a tricky (lit.: slippery) question'.

Some metaphors in this category focus on particular reasons for evaluating a person or another object as unreliable. For example, the Hungarian adjective *sikamlós*, originally meaning 'slippery', shifts to expressing the idea of vulgarity in *sikamlós téma* 'a vulgar (lit.: slippery) subject (especially one referring to sexual relations)', *sikamlós viccek* 'vulgar (lit.: slippery) jokes', see also Russian *skol'zkie frazy* 'vulgar phrases'. The Estonian lexeme *libe* 'slippery' metaphorically describes a flattering person as *libe inimene* whom it is difficult to believe, or the activities of such a person, in *libe jutt* 'a flattering (lit.: slippery) speech'.

Another class of metaphors stemming from 'slippery' is based on the idea of fast motion along a slippery surface. The metaphors of this SLIPPERY subdomain sometimes refer to **an action performed easily and quickly**. This metaphoric pattern is less frequent than the one with the meaning "non-trustworthy", but it is quite apparent in the case of the Finnish adjective *liukas* 'slippery'.[11] This lexeme is metaphorically used in *liukas varas* 'a skilled (lit.: slippery) thief', *liukas pelaaja* 'a skilled (lit.: slippery) player'; see also the uses of the derived adverb *liukkaasti* in (18) and (19). There are also some compound adjectives that include the component *liukas* with this metaphoric meaning, such as *liukasliikkeinen* 'agile, nimble (lit.: slippery + movements)' and *liukaskielinen* 'talkative (lit.: slippery + tongue)'.

FINNISH

(18) *Tiede+miehe-n   järki   pela-a   liukkaa-sti*
      science+man-GEN intellect play-3SG slippery-ADV
      'The scientist is very quick-witted (lit.: The scientist's intellect is playing in a slippery way)'

FINNISH

(19) *Auto-t liikku-vat liukkaa-sti   uude-lla   valtatie-llä*
      car-PL move-3PL slippery-ADV   new-ADESS   highway-ADESS
      'The cars are moving smoothly (lit.: in a slippery way) along the new highway'

---

11. This metaphor might seem to be closer to the idea of smoothness, see Section 3.2. However, the Finnish lexeme *liukas* belongs to the domain of SLIPPERY, but not to that of SMOOTH.

3.2   SMOOTH

The metaphors which involve lexemes meaning 'smooth' are mostly related to the **absence of defects or difficulties**, which follows from the prototypically positive evaluation of smooth surfaces and of their functionality. This usage embraces a wide range of situations, such as English *smooth speech, smooth transition, smooth implementation of the programme*, or Russian *Vse prošlo gladko* 'Everything went off smoothly'.

Some metaphors of this group express a slightly narrower meaning. Often they belong to the domain of speech and characterize either a person speaking confidently, or a fluent speaker of some language. Thus, Russian expressions *gladkaja reč'* (lit.: smooth speech) / *gladko govorit* (lit.: speaks smoothly) correspond to speaking confidently, while fluency is not present in Russian with any adjective close in meaning to 'smooth', but it is metaphorically described with an adjective *svobodnyj* or with an adverb *svobodno* (lit.: free / freely). An Udmurt metaphor *vol'yt veras'ke* (lit.: speaks smoothly) can refer to both frames depending on the broader context. In English *fluent* is defined as able to speak or write smoothly, easily or readily (cf. www.dictionary.com). In this and in many other examples, *smooth* is used as a synonym to *fluent*. In many other contexts *smooth* appears along with *confident*, *persuasive* – or is even used to render these meanings – as in *His smooth comments helped the committee to take the right decision*.

Another important metaphorical extension of SMOOTH, related to absence of defects, deals with human qualities. What may be focused on here is either a person's nice appearance or positive character traits. The first type of metaphor occurs with Estonian *sile*, as in *sileda näolapiga tüdruk* 'a girl with pretty (lit.: smooth) face', and Udmurt *vol'yt*, sometimes used in an idiomatic expression *s'artč'y kad' vol'yt* (lit.: smooth like a turnip) referring to a pretty girl with smooth outlines of the body. Positive character traits can be illustrated with Meadow Mari *jyvyža* developing the metaphorical meaning 'tender, soft' (about a person's character).

The concept of smoothness may call up metaphorical associations not only based on the absence of defects, but also based on **absence of distinctive features**. This becomes apparent in such examples as Spanish *tela lisa* 'plain (lit.: smooth) cloth' and *fachada lisa* 'a plain (lit.: smooth) façade' as in Spesivceva (2012: 32); Estonian *sile sõnastus* 'superficial, simplified (lit.: smooth) narrative'; and the Hungarian sentence in (20), where *sima* (lit.: smooth) means 'average, common'. Note that such metaphors often lack the positive evaluation typical of the previously discussed metaphorical shifts undergone by SMOOTH. What becomes more significant here is the caritive nature of SMOOTH which triggers the metaphor of something missing.

HUNGARIAN

(20) *Ez   egy   **sima**   bögre.*
     this INDEF smooth   cup
     'It's a common (lit.: smooth) cup.'

3.3   LEVEL

The semantic invariant of most metaphors developed in the LEVEL subdomain is the idea of **regularity or uniformity**. These metaphors may occur with different types of objects and therefore focus on different aspects of the basic invariant meaning.

Often lexemes meaning 'level' are metaphorically applied to static entities, as in English *evenly spaced desks*, *The wall is evenly covered with paint*, Russian *rovnyj zagar* 'an even suntan', *rovno rasstavit' stul'ja* 'to arrange chairs evenly', Finnish *tasainen rusketus* 'an even suntan', *tasainen väri* 'a regular colour'. A special case of this pattern takes place when a lexeme from the LEVEL subdomain modifies another qualitative lexeme and points to the steadiness of the quality. For example, this is the case of the English adverb *evenly* (*evenly red, evenly good quality*, etc.) and of Finnish *tasaisen*, which is the genitive of *tasainen* 'level', as in the following example:

FINNISH

(21) *Elämä on   **tasaise-n** tylsä*
     life   be.3SG level-GEN dull
     'Life is permanently dull'

Another metaphorical extension of LEVEL, which follows from its use with the names of static entities, expresses equality, cf. English *The score is even; Divide the dough into three even amounts,* or the Meadow Mari example in (22).

MEADOW MARI

(22) *Jerente məj deč-em   kugu-rak kap-an,   ijgot-šo   gəna məj*
     Jerente I   from-POSS1SG big-CMPR body-ATTR age-POSS3SG only I
     *den-em   **tör***
     with-POSS1SG level
     'Jerente is larger than me, but his age is equal to mine'   [Marlamuter]

In some languages such metaphors trigger further semantic extension of LEVEL terms into focus particles. This is the case for English *even*, German *eben*, and Russian *rovno*. However, this extension has not proved typical of the Uralic languages in our sample, but it has been widely discussed for Indo-European languages (cf. König 1991; Traugott 2006; Dobrovol'skij & Levontina 2012, and Luchina et al. 2013). Therefore, we do not discuss it here.

The idea of regularity / uniformity can be seen in the metaphors of LEVEL applied not only to static entities, but also to activities and processes. A good example is given by a Hungarian *egyenletes* 'level', which forms a wide range of such metaphors including *egyenletes lélegzés* 'even breathing', *egyenletes mozgás* 'uniform (lit.: level) motion', *egyenletes ritmus* 'steady (lit.: level) rhythm', *egyenletes sebesség* 'steady (lit.: level) speed', *egyenletes zúgás* 'steady (lit.: level) drone'. This pattern also exists outside the Uralic family, as in Russian *rovnoe dyxanie* 'even breathing', *rovnyj šag* 'even step', and the use of Korean *maekkulepta* 'level' in the contexts of steady voice or steady management of affairs (Ivanova 2011: 127). A closely similar type of metaphor involves use of LEVEL terms with reference to a calm person or their character / actions, as in Russian *rovnyj xarakter* 'a calm character', English *an even tone*, and the use of Chinese *píng* 'level' for describing a calm person (Kholkina 2014: 198).

Apart from the shift to the domain of regularity or uniformity, lexemes meaning 'level' have developed a second completely different metaphorical pattern. Based on the caritive component in the primary meaning of LEVEL, the metaphors within this pattern focus on **low intensity or absence of distinctive features**. A striking example is provided by Estonian *tasane* 'level', which expresses metaphorical meanings 'light' (pain, rain), 'slow' (motion, a river current), 'low, quiet' (sound, steps, waves). Similarly, Chinese *píng* 'level' follows a productive compounding pattern with the semantics of something ordinary or routine, cf. *píngcháng* 'ordinary, average (lit.: level + frequent)', *píngshí* 'ordinary, everyday (lit.: level + time)', *píngdàn* 'monotonous (e.g. about the style of a text; lit.: level + insipid)' (Kholkina 2014: 198–199). In English such metaphors are highly productive for the adjective *plain* (which is obviously related to the domain of surface texture, as its nominal use can refer to a large flat area of land): *a plain looking girl*, *plain food*, *plain English*, *plain text*, a *plainclothes police-officer*.

There are also highly distinctive uses of *flat*, *even* and *level* with metaphoric developments close to that shown for *plain*, but yet not exactly the same. *Flat*, in particular, has a broad combinability with many nouns to mean "dull, monotonous, lacking in expressiveness", for example: "*You go to work and you come at home in the evening, you watch this television… is a very **flat** life*" [COCA]; *flat character* as explained in https://www.thoughtco.com (> Humanities > Literature) as *a fictional character without any development or depth*; or 'fixed, not growing' with some other nouns as in: "*… problems led by online piracy that have resulted in falling or **flat** sales for five years*" [COCA]. However, there are a few nouns with which all three English adjectives – *flat, even* and *level* – can collocate, and one such noun is *tone*. A closer look at these combinations reveals the following subtle differences among the three adjectives (all examples are from COCA):

- *even* about someone's tone implies unprejudiced (*…in her judge's **even tone***), not showing unnecessary emotions (*He maintained his **even tone** and composed expression in spite of all provocation*); soothing or calming down (*Do not yell; speak in a low, **even tone** and show understanding* or *…to face the bear, talking to it quietly in a calm, **even tone** to let it know you are there*);
- *flat* used with *tone* refers to someone's intention to conceal something or not to let something out by one's way of speaking (*…she replied in a detached, **flat tone**. She was upset with me*); or even to communicate an impression contrary to what the words literally mean (*…with a **flat tone** that indicates that she might wish she could answer differently*; or *–You're a celebrity. –But there is nothing congratulatory in the **flat tone** of his voice*);
- *level* is much less frequent in combination with *tone* than *even* and *flat*. *Level* renders the meaning of a deliberately chosen or strictly controlled tone of voice (*I asked in a deliberately **level tone***; *Wilder paused until he could be sure of a **level tone** of his voice*).

3.4   ROUGH

Lexemes denoting roughness typically develop metaphors **involving defects or difficulties**. Often they describe the lack of precision or poor quality of some action, as in English *a rough estimate, a rough draft*, Russian *grubo skoločennyj stol* 'a table crudely knocked together', Erzya *kaz'amo ez'em* 'crude bench'. This metaphorical pattern is seen in the use of ROUGH for describing illiterate speech lacking in confidence, cf. Udmurt *kylyz šakyres* 'His language is poor (lit.: his tongue / language is rough)'.

It is often the case that adjectives from the ROUGH domain develop metaphors of human qualities or actions indicating that a particular person is difficult to deal with. Most often these metaphors refer to impoliteness, such as Russian *grubyj čelovek* 'a rude person', *grubye slova* 'rude words', English *coarse joke, coarse language*, Erzya *kaz'amo loman'* 'a rude person', Estonian *krobelised kombed* 'coarse manners'. Sometimes ROUGH, if applied to human beings, describes a strict or severe person, like in Spanish *hombre áspero* 'a severe (lit.: rough) person' (Spesivceva 2012: 54) or in Example (23) from Estonian.

ESTONIAN
(23) *Isa   ol-i   laste   vastu   **kare***
      father be-PST child:GEN.PL with   rough
      'A father was strict with his children'

Some metaphors involving ROUGH refer to unpleasant physiological sensations. There are some expressions with these lexemes denoting an unhealthy person or their body parts; see English *to feel rough,* the Estonian expression in (24), and the Tundra Nenets expression in (25).

ESTONIAN

(24) *Kurk   on   külmetuse-st **kare***
       throat be.PRS.3 chill-EL   rough
       'He has caught a cold and has a sore throat (lit.: His throat is rough because of
       the chill)'   [EVS]

NENETS

(25) *sæw-myih   / sæw-xøna-nyih   **nasorta**.*
       eye-POSS1SG   eye-LOC-POSS1SG be.rough:3SG
       'I have a speck in my eye (lit.: My eye is rough / It is rough in my eye)'

Similarly, ROUGH metaphorically describes qualities unpleasant for sense organs. Such metaphors have been attested in the domains of taste (English *rough wine*, Spanish *sabor áspero* 'an astringent taste') and sound (English *The clutch sounds rough – better get it checked*). As regards unpleasant senses, metaphorical uses of ROUGH are often related to a hoarse human voice. Interestingly, in different languages these metaphors may indicate different reasons for a voice being hoarse. Thus, an Estonian metaphor *kare haal* 'a hoarse (lit.: rough) voice' describes a voice hoarse for any reason (chill, screaming, a natural quality), as does its Erzya counterpart *kaz'amo vajgel'*. The Hungarian expression *érdes hang* 'a hoarse (lit.: rough) voice' describes only a voice which is always hoarse and is not applicable if a voice has become hoarse as a result of shouting or a sore throat.

Finally, lexemes with the original meaning 'rough' can metaphorically characterize unfavourable conditions. For instance, English *rough* has a vast list of such collocations, e.g. *a rough night, a rough journey, a rough day, rough going*. A variant of this strategy is represented by the metaphors describing bad weather conditions, such as Estonian *kare põhjatuul* 'piercing (lit.: rough) northern winds', *Siberi kare talv, kliima* 'the severe Siberian winter, climate', Shoksha Erzya *kaz'ama varma* 'biting (lit.: rough) wind', *kaz'ama lov* 'biting (lit.: rough) snow (during a snowfall)', and *kaz'ama t'el'is* 'severe (lit.: rough) winter'.

Some metaphors of the ROUGH domain follow a completely different pattern, focusing not on defects or difficulties, but on **the intensity of an action or of a quality**. This metaphorical pattern seems to stem from applying physical force with the use of rough objects (such as, for example, abrasive paper or abrasive brick). An impressive example is provided by Northern Udmurt ideophones with the root *čaž-*, describing rough surfaces which derive a metaphorical meaning 'refreshing,

quenching one's thirst' (about a drink). According to the interpretations of native speakers, this metaphor is motivated by the fact that such drinks can cause a burning sensation in one's throat, i.e. the semantic shift is linked here to the impact of something rough that can be felt on an object.

Lexemes from the source domain of roughness can serve as intensifiers in quite a few abstract contexts. This can be observed in the case of English *rugged / ruggedly*; consider such examples as *The camera combines **rugged** reliability with unequalled optical performance and speed* [WordBanks]; *The telescope is **ruggedly** solid with nothing that can be easily damaged* [COCA]; *… facing the **ruggedly** competitive conditions in a country whose population has risen from 300 million in 1960 to 1 billion today* [COCA]; *… galvanizing interpretations of these **ruggedly** intense, expansive and unapologetically romantic compositions…* [COCA]. Russian *grubaja ošibka* (lit.: a coarse mistake) denotes a blunder, and similar metaphors have been attested in German (Bons 2009: 306–307), such as *grobe Mängel* 'grave (lit.: coarse) defects', *grob ungerecht* 'very (lit.: coarsely) unfair'. In colloquial Hungarian there is a productive pattern of using an adjective *durva* 'coarse' (or an adverb *durván* 'coarsely') as an intensifier, either with positive or with negative connotations; see *durván elfáradtam* 'I am terribly (lit.: coarsely) tired', *durván megijedtem* 'I was terribly (lit.: coarsely) frightened', *durva autó* 'a cool (lit.: coarse) car', *durván szeretem* 'I am madly (lit.: coarsely) in love'.

4. Discussion

The analysis of the literal and metaphoric uses of texture expressions allows us now to draw the conclusions and make some theoretical implications.

The first of these conclusions deals with the organization of the **sensory lexicon** domain. How language categorises experiences of the five senses is a popular issue in cognitive semantics (Viberg 1983; Majid & Levinson 2011; Levinson & Majid 2014; Koptjevskaja-Tamm 2015). In our discussion of the frames included in the domain of surface texture (see Section 2), we showed that one of the most significant parameters in their cross-linguistic categorisation is the way a surface is perceived (perception is especially relevant for lexemes describing absence of roughness). From the physiological point of view, there are two channels of perception prototypically operating with surface texture: visual and tactile. However, linguistically these frames are much more subtly categorised. There is a specific type of surface perceived by touch, namely, slippery surfaces with their inherent negative functionality usually leading to their special encoding in languages. The frames of visually perceived surfaces are not a homogenous phenomenon either:

they are subdivided into artefacts and land areas, the latter being further classified into roads, water surfaces, and landscape without mountains. Sometimes tactile and visual features interplay in the semantics of a lexeme, as is the case with the frame of sleek surfaces (both smooth to the touch and reflecting light).

Another interesting point concerns **antonymic relations** in the lexicon. Antonyms have often been regarded as asymmetrical in their semantic features and combinability in semantics (see, among others, Apresjan 1974; Cruse 1986; Croft & Cruse 2004). But this phenomenon has not been thoroughly investigated in linguistic typology. Thus, it is not quite clear what semantic entities are most commonly involved in this asymmetry cross-linguistically. Our data include two antonymic zones: absence of roughness (SLIPPERY, SMOOTH, LEVEL) and roughness. The literal uses of the three subdomains are subcategorised in quite different ways. Their patterns of metaphoric shifts, described in Section 3, are not fully symmetrical either. A possible explanation is that frames of roughness have an additional element (roughness itself), which triggers new semantic oppositions in the literal uses and new grounds for semantic shifts. At the same time, the caritive zones of SLIPPERY, SMOOTH and LEVEL lack this semantic element, and the main focus in their categorisation is on the type of surface per se (rather than on the type of external elements on it). A further search for the patterns of semantic asymmetry reproduced in various lexical domains is a challenge for lexical typology.

Finally, our research contributes to the issue of **language sampling**, which is highly important for linguistic typology (considered in Croft 1990; WALS; Bakker 2010, among others). According to the traditional view, a sample must be representative, which means it must include languages from different families and areas, so that the research can fully embrace the cross-linguistic diversity.

There is, however, another approach to typological studies, called intragenetic typology. As argued by Alexander Kibrik in (Kibrik 1998, 2003, 2009), this approach, when applied to grammar studies, first, provides more subtle cross-linguistic differences, which might be difficult to notice when working with a broader sample; second, it provides more systematic data for diachronic typology; third, it can serve as a starting point for a broader study based on the comparison of more diverse language groups.

The issue of language sampling has also been raised with respect to lexical typology. Some authors have pointed out that studying a lexical group in closely related languages can reveal a considerable number of semantic oppositions: D'urovič (2000) on verbs of cutting and breaking in Russian and Slovak; Rakhilina & Prokofieva (2004, 2005) on verbs of rotation and oscillation in Russian and Polish; and Majid et al. (2007) on verbs of cutting and breaking in English, German,

Swedish, and Dutch.[12] The sample in our paper is a bit different. While previous research usually focused on 2–4 languages belonging to the same family to show that the structure of semantic domains in closely related languages is not identical, our research covers more languages from the same family (10 Uralic languages), as well as 5 languages from other families in order to check the typological validity of the conclusions drawn in the intragenetic study.[13] The most important conclusions made on the basis of our sample are the following.

There are many cognates (words in related languages having a common ancestor and showing a regular phonetic correspondence), which is an important point in intragenetic studies in lexical semantics. Among those are two interesting cognate sets in the ROUGH domain: Khanty *karəŋ*, Hungarian *kérges* (set 1), and Estonian *kare*, Finnish *karhea* and *karkea* (set 2; see (UEW: 148), (SSA I: 314) for the etymological data). The adjectives forming these sets have completely different semantic scopes. Native speakers of Khanty often link *karəŋ* to a noun *kar* 'crust, bark', and what is prototypically described by the adjective *karəŋ* is surfaces covered with some crust. Its Hungarian cognate *kérges* is also transparently related to a noun denoting crust (*kéreg*), but it shifts to the frame of large roughness (while a dominant adjective for ROUGH is *érdes*). As regards the cognates from Estonian and Finnish, the Estonian adjective *kare* describes a long list of rough surfaces (a cat's tongue, skin, unshaven cheeks, rough paper in old notebooks, etc.) and productively develops metaphors (hoarse voice, severe winter, strict or severe person). Its Finnish cognates, *karhea* and *karkea*, differ in the size of irregularities they describe, and develop some metaphors not equal to those of their Estonian cognate from set 2 (e.g. *karkeat sanat* 'rude words', *karkea arvio* 'a rough estimate'). Examples of this kind provide evidence of different stages of diachronic development in the lexicon.

Interestingly, there are some cognate sets used approximately in the same way across the languages in their literal meanings, but demonstrating completely different metaphoric patterns. For example, Finnish *tasainen* and Estonian *tasane,* whose source meanings are both dominant lexemes over the frames of level surfaces, have two completely different metaphoric developments. Finnish *tasainen* refers to regularity or uniformity (an even suntan, evenly spaced chairs, steady motion), while

---

12. Another important area of research focusing on lexicons of closely related languages deals with semantic reconstruction (Dybo 1996, 2006). However, the tasks set in our paper are a bit different and mostly concern the synchronic lexical typology.

13. There is also a paper (Majid et al. 2015) which presents an exception, dealing with the sample of 12 Germanic languages. It is discussed further, but at the same time its main research goals are different from ours. See also Koptjevskaja-Tamm (this volume) for the analysis of temperature terms in a group of related (Slavic) languages.

Estonian *tasane* indicates low intensity or absence of distinctive features (a modest person, a low sound, slight pain, slow motion).

Examples of this kind and the data laid out in Sections 2 and 3 indicate clearly that the Uralic languages show substantial cross-linguistic variation in how the domain of surface texture is organized. It is interesting to compare our results with those of Majid et al. (2015), who compared the stability of lexemes in four different domains (colours, body parts, containers, spatial relations) across 12 closely related Germanic languages. They argue that colour systems are quite stable in both their form and in their semantics within the sample analysed. However, our data on another group of quality terms differ: the textural lexicon appears to show more prominent variation across a sample of related languages. The reasons for such a result are not clear; they are a challenge for future research on the intragenetic typology of qualities.

At the same time, our comparison of the Uralic data with materials of the other five languages in the sample has revealed that the semantic oppositions between the literal meanings tend to be reproduced outside the Uralic family, as do the basic metaphoric patterns. Therefore, the lexico-typological conclusions drawn for the Uralic family appear valid from a broader typological perspective. This assertion surely needs additional confirmation from other semantic domains, but at least it is clear that studying closely related languages can serve as a sound basis for research in lexical typology.

Abbreviations

1, 2, 3	the 1st, 2nd, 3rd person	IPFV	imperfective
ADESS	adessive	LOC	locative
ADV	adverb	M	masculine
ATTR	attributive	NMN	nominalization
CMPR	comparative	NOM	nominative
COM	comitative	NPST	non-past
CVB	converb	O	object conjugation
DAT	dative	OBL	oblique
DEF	definite	PASS	passive
DIM	diminutive	PL	plural
EL	elative	POSS	possessive
F	feminine	PRS	present
GEN	genitive	PST	past
IMP	imperative	PST2	the 2nd past
INDEF	indefinite	SG	singular.
INF	infinitive		

References

Apresjan, Ju. D. 1974. *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka* [Lexical Semantics. Synonymic Tools of Language] Moscow: Nauka.

Bakker, D. 2010. Language sampling. In *The Oxford Handbook of Linguistic Typology*, J. J. Song (ed), 100–127. Oxford: Oxford University Press.

Bons, I. 2009. *Polysemie und Distribution Zur Theorie und Methode einer korpusbasierten Semantik deutscher Adjektive*. Gießen: Gießener Elektronische Bibliothek.

Croft, W. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

Croft, W. & Cruse, D. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803864

Cruse, D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Djurovič, L. 2000. Soblazn rodstva (o značenii leksem v rodstvennyh jazykah) [Temptation of relatedness (on the meanings of lexemes in genetically related languages)] In *Slovo v tekste i slovare. Sb. statej k 70-letiju akad. Ju.D. Apresjana* [A Word in a Text and in a Dictionary. To the 70-th Anniversary of Ju.D. Apresjan], L. L. Iomdin & L. P. Krysin (eds), 77–81. Moscow: Jazyki russkoj kultury.

Dobrovoľskij, D. O. & Levontina, I. B. 2012. O sinonimii fokusirujuščih častic (na materiale nemeckogo i russkogo jazykov) [Synonymous focus particles in German and Russian]. In *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" -2012*, 138–149. Moscow: RGGU.

Dybo, A. V. 1996. *Semantičeskaja rekonstrukcija v altajskoj etimologii. Somatičeskije terminy (plečevoj pojas)* [Semantic Reconstruction in Altaic Etymology. Somatic Terms (Shoulder Girdle)]. Moscow: Jazyki slavjanskoj kuľtury.

Dybo, A. V. 2006. Indoeuropeans and Altaians through the Linguistic Reconstruction. In *Kinship in the Altaic World. Proceedings of the 48th PIAC*, E. V. Boykova & R. R. Rybakov (eds), 88–102. Wiesbaden: Harrassowitz Verlag.

Haspelmath, M. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison In *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, Vol. 2, M. Tomasello (ed), 211–242. Mahwah, NJ: L. Erlbaum Associates.

Ivanova, E. Yu. 2011. Netriviaľnye svojstva poverhnostej kak topologičeskih ob''ektov: materialy k leksičeskoj tipologii (russkie i korejskie prilagateľnye) [Non-trivial properties of surfaces as topological objects: data for lexical typology (Russian and Korean adjectives)]. MA thesis, Russian State University for the Humanities.

Kashkin, E. V. 2011a. Prilagateľnye, opisyvajuščie svojstva poverhnostej, v ižemskom dialekte komi-zyrjanskogo jazyka (sopostaviteľno-tipologičeskoe issledovanie) [Adjectives denoting surface texture in Izhma Komi: a cross-linguistic perspective]. *Acta Linguistica Petropolitan* VII(3): 302–307.

Kashkin, E. V. 2011b. Na puti k kompjuternomu analizu i sintezu semantiki: opyt issledovanija erzjanskih i udmurtskih prilagateľnyh, opisyvajuščih fakturu poverhnostej [Towards a computational semantic analysis and synthesis: some remarks on Erzya and Udmurt adjectives of surface texture]. In *Problemy kompjuternoj lingvistiki* [Issues in Computational Linguistics], Vol. 5, A. A. Kretov (ed.), 137–153. Voronezh: Voronezh State University.

Kashkin, E. V. 2012. Kategorizacija faktury poverhnostej v zapadnyh govorah hantyjskogo jazyka: sopostaviteľno-tipologičeskoe issledovanie [The categorisation of surface texture in Western Khanty: a cross-linguistic perspective]. *Ural-Altaic Studies* 2(7): 26–38.

Kashkin, E. V. 2013a. Jazykovaja kategorizatsija faktury poverhnostej (tipologičeskoe issledovanie naimenovanij kačestvennyh priznakov v ural'skih jazykah) [Language categorisation of surface texture (a cross-linguistic study of quality expressions in Uralic languages)]. Candidate's dissertation in philology, Moscow State University.

Kashkin, E. V. 2013b. K leksičeskoj tipologii priznakovyh slov, opisyvajuščih fakturu poverhnostej: dannye vengerskogo jazyka [Towards a lexical typology of quality expressions denoting surface texture: evidence from Hungarian]. In *Problemy leksiko-semantičeskoj tipologii* [Issues in Lexico-Semantic Typology], Vol. 2, A. A. Kretov (ed), 113–140. Voronezh: Voronezh State University.

Kholkina, L. S. 2014. Kačestvennye priznaki v kitajskoj leksike: opyt tipologičeskogo opisanija [Qualitative concepts in Chinese vocabulary: a typological description]. Candidate's dissertation in philology. Moscow, Moscow State University.

Kibrik, A. E. 1998. Does intragenetic typology make sense? In *Sprache in Raum und Zeit: Beiträge zur empirischen Sprachwissenshaft*, W. Boeder, Ch. Schroeder, K. H. Wagner & W. Wildgen (eds), 61–68. Tübingen: Gunter Narr Verlag.

Kibrik, A. E. 2003. Rodstvennye jazyki kak ob''ekt tipologii [Genetically related languages as an object of typology] In *Konstanty i peremennye jazyka* [Constants and Variables of Language], A. E. Kibrik, 191–195. St. Petersburg: Aleteya.

Kibrik, A. E. 2009. In defence of intragenetic typology: evidence from clause coordination in Daghestanian languages. In *Abstracts of the Eighth Biennial Conference of the Association for Linguistic Typology*, <http://lsa2009.berkeley.edu/alt8/kibrik_aleksandr_abstract.pdf> (1 October 2019).

Koptjevskaja-Tamm, M. (ed). 2015. *The Linguistics of Temperature*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.107

König, E. 1991. *The Meaning of Focus Particles. A Comparative Perspective*. London: Routledge.

Koshkareva, N. B., Kashkin, E. V., Koryakov, Yu. B., Kazakevich, O. A., Burkova, S. I., Muravyev, N. A. & Budyanskaya, E. M. 2017. *Dialektologičeskij atlas ural'skih jazykov, rasprostranennyh na territorii Jamalo-Nenetskogo avtonomnogo okruga* [Dialectological Atlas of the Uralic Languages Spoken in the Yamalo-Nenets Autonomous District]. Kaliningrad: ROS-DOAFK.

Lederman, S. J. & Klatzky, R. L. 2009. Haptic perception: a tutorial. *Attention, Perception, & Psychophysics* 71 (7): 1439–1459. https://doi.org/10.3758/APP.71.7.1439

Levinson, S. C. & Majid, A. 2014. Differential ineffability and the senses. *Mind & Language* 29: 407–427. https://doi.org/10.1111/mila.12057

Luchina, E. S., Reznikova, T. I. & Stenin, I. A. 2013. Atributivy kak istočnik grammatikalizacii: 'prjamoj' i 'rovnyj' v russkom, nemeckom i finskom jazykah [Attributives as a source of grammaticalization: 'straight' and 'level' in Russian, German and Finnish]. In *Tipología léxica*, R. Guzmán Tirado & I. Votyakova (eds), 123–129. Granada: Jizo Ediciones.

Majid, A., Gullberg, M., van Staden, M. & Bowerman, M. 2007. How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics* 18(2): 179–194. https://doi.org/10.1515/COG.2007.007

Majid, A., Levinson, S. C. (eds). 2011. The senses in language and culture. *The Senses & Society* 6(1): 5–18. https://doi.org/10.2752/174589311X12893982233551

Majid, A., Jordan, F. & Dunn, M. 2015. Semantic systems in closely related languages. *Language Sciences* 49: 1–18. https://doi.org/10.1016/j.langsci.2014.11.002

Rakhilina, E. V. & Prokofieva, I. A. 2004. Rodstvennye jazyki kak ob''ekt leksičeskoj tipologii: russkie i pol'skie glagoly vraščenija [Genetically related languages as an object of lexical typology: Russian and Polish verbs of rotation]. *Voprosy jazykoznanija* 1: 60–78.

Rakhilina, E. V. & Prokofieva, I. A. 2005. Russkie i pol'skie glagoly kolebateľnogo dviženija: semantika i tipologija [Russian and Polish verbs of oscillation: semantics and typology]. In *Jazyk. Ličnosť. Tekst. Sb. k 70-letiju T.M. Nikolaevoj* [Language. Personality. Text. To the 70th anniversary of T.M. Nikolaeva], V. N. Toporov (ed), 304–314. Moscow: Jazyki slavjanskih kuľtur.

Spesivceva, A. A. 2012. Semantika prilagateľnyh, opisyvajuščih kačestva poverhnosti, i prilagateľnyh so značeniem "pustoj" i "polnyj" (na materiale ispanskogo jazyka) [The semantics of adjectives denoting surface properties and adjectives 'full' – 'empty' (evidence from Spanish)]. MA thesis, Moscow, Russian State University for Humanities.

Traugott, E. 2006. The semantic development of scalar focus modifiers. In *The handbook of the history of English*, A. van Kemenade, B. Los (eds), 335–359. Malden, Oxford, Carlton: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757048.ch14

Viberg, A. 1983. The Verbs of Perception: A Typological Study. *Linguistics* 21(1): 123–162. https://doi.org/10.1515/ling.1983.21.1.123

Vinogradova, O. I. 2013. K leksičeskoj tipologii priznakovyh slov, opisyvajuščih fakturu poverhnostej: dannye anglijskogo jazyka [Towards a lexical typology of surface texture expressions: evidence from English]. In *Problemy leksiko-semantičeskoj tipologii* [Issues in Lexico-Semantic Typology], A. A. Kretov (ed), 39–72. Voronezh: Voronezh State University.

Sources

[BNC] – British National Corpus, <https://www.english-corpora.org/bnc/> (1 October 2019).

Buzakova, R. N. 1982. *Slovar' sinonimov erz'anskogo jazyka* [The dictionary of synonyms in Erzya]. Saransk: Mordovskoe knižnoe izdateľstvo.

[COCA] – The corpus of contemporary American English, <https://www.english-corpora.org/coca/> (1 October 2019).

[EVS] – *Eesti-vene sõnaraamat* [Estonian-Russian Dictionary], <www.eki.ee/dict/evs> (1 October 2019).

Macmillan English Dictionary for Advanced Learners. International Student Edition. 2006. Oxford: Macmillan Education.

[Marlamuter] – Marlamuter: an online dictionary of Mari, <http://marlamuter.com/> (27 April 2018).

[RNC] – The Russian National Corpus, <www.ruscorpora.ru> (27 April 2018).

SSA, 1992–2000. *Suomen Sanojen Alkuperä: Etymologinen Sanakirja*. Helsinki: Suomalaisen Kirjallisuuden, Seura: Kotimaisten kielten tutkimuskeskus.

UEW – Rédei, K. 1988. *Uralisches Etymologisches Wörterbuch. Band 1*. Budapest: Akadémiai Kiadó.

[WALS] – Dryer, M. S. & Haspelmath, M. (eds). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, <http://wals.info/> (27 April 2018).

A new approach to OLD studies

Anastasia Vyrenkova[1], Ekaterina Rakhilina[1,2]
and Boris Orekhov[1]

[1]HSE University, Moscow / [2]Vinogradov Russian Language Institute
of the Russian Academy of Sciences

This chapter gives a description of the OLD terms typology across 78 world languages. The sample size in the research is relatively big, as compared to the other studies in this volume, which is attributable to two main reasons: the semantics of OLD terms has already been thoroughly described for at least two languages – English and Russian – and the lexemes with this meaning are in most cases featured by bilingual dictionaries. On the basis of the dictionary data supported by additional expertise we propose a semantic map of the OLD domain. The paper also outlines some cross-linguistically recurrent colexification patterns and addresses the specificity of homonymous combinations of OLD attributes with different semantic classes of nouns.

Keywords: qualitative adjectives, semantics of oldness, lexical typology, semantic maps

1. Introduction

This paper is both similar to and dissimilar from the other studies included in the current volume. Just as the other research presented in this book, we draw upon the frame-based approach to lexical typology (see Rakhilina & Reznikova 2016 and Chapter 1 in this volume). While presenting opportunities for an accurate and detailed cross-linguistic analysis of lexical systems, the frame-based approach is, however, quite time-consuming and resource-demanding, which inevitably leads to reducing the language sample to ten or fewer languages.

In the present study, we made an attempt to overcome this constraint typical of a frame-based project, and to expand the scope of our sample. To make such an endeavor possible, we picked a domain that (a) has been thoroughly investigated in at least one language, and (b) is adequately presented in dictionaries of multiple languages. Reliance on prior research provides the basis for extracting frames that

constitute the domain, thus saving us the time that would otherwise be spent on the meticulous and painstaking analysis of word combinations in the researcher's native language, while using dictionary data enables significant expansion of the sample.

The present research focuses on the semantic domain of the property OLD which meets both of the above requirements. This domain has already been investigated in Russian and English (Beard 1991; Taylor 1992; Rakhilina 1999), and the lexemes that express its meanings are in most cases featured by bilingual dictionaries,[1] which allowed us to expand the sample to 78 genetically different languages. Apart from a typological analysis of the OLD domain on the basis of extended cross-linguistic data, this paper approaches two more issues: describing the relations between the lexemes that constitute the domain, and examining the finer-grained lexical oppositions that express the OLD meanings in individual languages of the sample.

The paper is organized in the following way. In Section 2, we analyze the structure of the domain under examination; Section 3 describes the methodology and the data; and Section 4 reports the typological results. Finally, Section 5 outlines some considerations on the dominant systems (see Section 2), and Section 6 focuses on specific features of various lexical systems representing the OLD domain in the world languages.

2.   Semantics of OLD

We begin by observing the structure of OLD as a semantic domain. The semantics of oldness has been a subject of several theoretical studies (cf. Dahl 1970; Beard 1991; Taylor 1992; Rakhilina 1999). Beard (1991) distinguishes between the two situations that are essential for the domain: 'being old in age' (e.g. *old man*) and 'being involved in an old friendship' (e.g. *old friend*). Taylor (1992) proposes a more elaborate structure consisting of three situations: 'smth/smb of long standing' (e.g. *old friend*), 'former smth/smb' (e.g. *old girlfriend*), 'smth which no longer exists' (e.g. *old regime*). In the current paper, we follow Rakhilina (1999), who established four typical situations associated with oldness (see below).

In her study, Rakhilina (1999) focuses on Russian nominal constructions containing the adjective *staryj* 'old', in which the nouns receive four different interpretations; for convenience of demonstration, we will indicate them by the following

---

1.   The list of dictionaries used in this study is presented at http://nevmenandr.net/pages/old.php.

labels (which correspond to the typical situations or, following the MLexT terminology, the frames, see Chapter 1 in this volume):

1. 'old person'
2. 'old clothes'
3. 'old boss'
4. 'old coin'

The first frame ('old person') illustrates the situation of gradual change of an object over time. This interpretation can be defined as 'an object that came into being a long time ago and has existed for a long time, which has led to its gradual change'; it is applicable to people's age as well as to gradual ageing of some natural objects such as trees or mountains.

The second frame ('old clothes') applies to artifacts that have a fixed lifespan or typically wear out after having been in use for a long time; these include, among the others, clothes and buildings. Note that artifacts which become more valuable in the course of time belong to frame 4: thus, a phrase like *old clothes* corresponds to the second frame only if it names 'old items that are worse than new ones' (i.e. they are ramshackle and shabby).

The third frame ('old boss') characterizes changeable statuses, e.g. *previous husband* in the sense of 'the one who used to be smb's husband'; *former house* in the sense of 'the one smb used to live in'; *former hairdo* in the sense of 'the one I used to typically wear', etc. There is a certain similarity between this frame and the previous one: while Frame 2 denotes approaching of the limit of an object's lifespan, Frame 3 implies the limitedness of a status. However, for the latter, the notion of boundedness is not the central component – instead, Frame 3 focuses on replacing an 'old' object with another ('new') one.

The fourth frame ('old coin') is generally applicable to objects that were created long ago but have not lost their value or have even increased in it. This is true of "creative objects" (like authored pieces of art belonging to a certain epoch, e.g. *old painting*, *old icon*) as well as objects representing a certain past/historical period (e.g. *old town*, *old coin*).

In what follows we address the typological regularities attested in the domain of OLD by analyzing the lexicalization patterns for these four frames. The next section will give a brief outline of the collected data and the methodology which enabled us to uncover these regularities.

3.   Data

To conduct the present research, we collected lexical data from 78 languages from 11 families, using bilingual dictionaries as the primary source. The analysis consisted in examining dictionaries that have Russian or English as the source language and comparing the obtained entries with the corresponding entries in dictionaries with Russian or English (respectively) as the target language. We started our analysis with the lexeme meaning 'old' in the situations matching the first of the frames presented in Section 2, and then checked if that word could be legitimately used for the other three frames. It turned out that the dictionaries vary in their approaches to presenting the meanings – some of them organize lexical entries in such a way that all the four frames are easily detectable (cf. the dictionaries of Spanish, French, Kalmyk, and Vietnamese), while others provide indiscriminately generic definitions of lexical items (for example, the dictionary of Bashkir).[2]

Considering the heterogeneity of the dictionary data, we went on to verify the results with the help of experts – linguists and native speakers – at the second step of the research. This was done with the use of corpora and questionnaires that resulted in a clearer image of each frame. Some of the sample questions from the typological questionnaire are presented below (the full questionnaire is available in the appendix to this chapter):

1.   [Old] people get cold easily.
2.   [Old] women look younger than old men.
3.   His [old] wife was kind-hearted, and the new one is beautiful but bad-tempered.
4.   Remarkable works of art by [Old] Masters are kept in museums.
5.   There is an [old] oak close to our house; it is probably 200 years old.
6.   We have an [old] raincoat in the closet, it's time to throw it away because it leaks.

4.   Typological results

The structure of the OLD domain in world languages suggests several options for frame colexification strategies: there are dominant lexical systems, where a single lexeme covers all the four frames, distributive systems, in which each frame is accommodated with a lexical unit which is designated to it and cannot be used to

---

2.   References to the dictionaries of each language are available at http://lcph.bashedu.ru/cgi-bin/oldmap.html.

describe other frames, and also interim systems with two or three lexemes comprising the domain.

All of the options described above have been attested in the sample; dictionary analysis and the subsequent evaluation by experts revealed the following colexification patterns:

a. all the four frames are accommodated with at least one lexeme (dominant system; Slavic: Russian *staryj*, Polish *stary*, Bulgarian, Serbian *star*, etc., Baltic: Lithuanian *senas*, Latvian *vecs*; Germanic: English *old*, German *alt*, Danish *gammel*; Romance: Latin *vetus*, Spanish *viejo*, Italian *vecchio*, Romanian *vechi*, and Finno-Ugric: Estonian *vana*);

b. the four frames are covered with at least two lexemes; one is used to describe human age (and the age of personalized objects), i.e. 'old person', while the other one is to collectively denote the rest of the frames (binary system, type 1; e.g. Modern Greek, Albanian, Hindi, Tamil, Turkish, and Uzbek);

c. the four frames are covered with at least two lexemes, where one lexeme describes cyclic positions and states ('old boss'), while the other applies to the remaining frames (binary system, type 2; e.g., Icelandic and Finno-Ugric – Finnish, Hungarian, etc.);

d. the four frames are covered with at least three lexemes: the first lexeme is used exclusively to describe human age (and the age of personalized objects, i.e. the 'old person' frame), the second lexeme denotes artifacts with a fixed lifespan ('old clothes'), and the third one is designated for the 'old boss' and the 'old coins' frames (distributive system, type 1; Vietnamese, Turkic – Bashkir, Kirghiz, Crimean Tatar, and Turkmen);

e. each frame is expressed with a dedicated lexeme exclusive to it (distributive system, type 2, attested in Quechua, Nanai, Komi, Malagasy, Ossetian, Shugni, and Georgian).
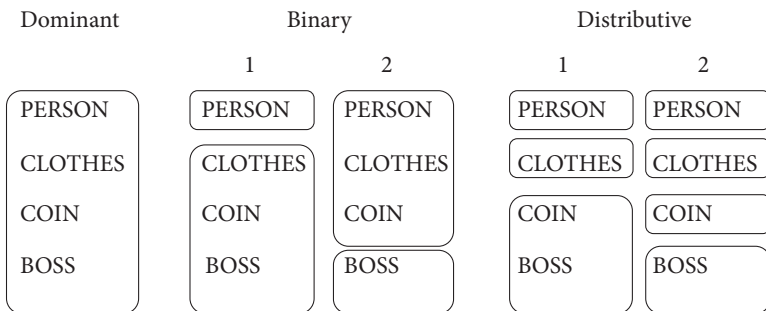
Figure 1. Semantic map of the OLD domain

Figure 1 illustrates these patterns as a semantic map; the curved closed lines indicate the boundaries of a lexeme in the domain – e.g., in the dominant system, one lexeme expresses all the four frames. In accordance with the semantic map connectivity hypothesis, which posits that the nodes should be located in a way that predicts language-specific combinations of meanings (Haspelmath 2003; Croft 2001; for the review of the semantic map model see also Georgakopoulos & Polis 2018 and Rakhilina et al., in press), we arranged the frames in the following order: 'old person' – 'old clothes' – 'old coins' – 'old boss'. Thus, we conjecture that it is highly unlikely for any language to lexically oppose 'old person' and 'old coins', on the one hand, to 'old clothes' and 'old boss', on the other – as well as to juxtapose 'old person' and 'old boss' to 'old clothes' and 'old coins'.

The results also show that typological distribution of the OLD terms is not arbitrary and is based on three features. The first one is **anthropocentricity**. As in many other domains, persons and personalized objects are cognitively salient. Consequently, there are languages that emphasize the prominence of the 'old person' frame and express human age with a dedicated lexeme (cf. Japanese *oita* 'old of man/animate being', Ossetian *zaerond* 'old of man', Hindi *boodha* 'old of man'). The second one is **positive/negative evaluation**. We have noted above that the second frame for clothes and other artifacts with a limited lifespan is only applicable to constructions in which the old items are interpreted as inferior to new ones – i.e. useless and unpractical. This is the only frame for which negative connotation is considered to be an intrinsic feature, and this feature becomes salient enough to grant 'old clothes' a separate lexeme in a number of languages, cf. Mandarin Chinese 老 *lǎo* 'aged/experienced/ancient' vs. 旧 *jiù* 'neg: having lost its functionality'. The final feature that supposedly affects lexicalization patterns across languages is **cyclicity**. The third frame ('old boss') is semantically opposed to the others in that it covers a special class of nouns expressing temporary statuses or positions, as contrasted with 'old person', 'old clothes', and 'old coins', which refer to an intrinsic property of an object. This frame can also be lexicalized with a dedicated unit, cf. Finnish *entinen* 'previous/former,' Udmurt *az'lo* 'previous/former,' Swedish *före* 'previous/former'.

The overall data analysis may be concluded into the following outcomes.

1. One of the polar systems – the distributive system with four lexical members each corresponding to a separate frame – is the rarest type found in our sample (attested in 7 languages).
2. The distributive system of three lexemes has been found in 25 languages. Apart from marking 'old person' and 'old clothes' with a special lexicalization means, it displays a unification of two other frames – 'old boss' and 'old coins'. This fact has a straightforward explanation, since both frames share a common feature – they highlight the object's relatedness to the previous period: 'old boss' to a period of the previous status and 'old coins' to an expired epoch to which the object belongs.

3. There are two subtypes among the binary systems: anthropocentric ('old person' vs. the remaining frames) and cyclicity-oriented ('old boss' vs. the remaining frames). Although there is also a possibility to oppose 'old person' and 'old clothes' (as related to positive and negative sides of the ageing period) to 'old coins' and 'old boss' (as pointing to the earlier temporal stretch), such lexical system occurred only once within the sample, in Russian Sign Language.
4. The final type under scrutiny, and a seemingly simple one, is the dominant lexical system, in which only one lexeme dominates across the domain. It has to be noted, however, that the picture is more nuanced, since one word operating in all the frames does not mean that it is the only lexical means to express the concept of oldness.

Dominant systems will be detailed in Section 5.1, where we look at the polysemy of noun combinations with the dominant *old* lexeme, and in Section 5.2, which focuses on the rules of lexical distribution for the dominant lexeme and its quasi-synonyms.

5. Dominant systems

5.1 Polysemy across frames

In lexical systems where a single OLD lexeme may be used to express all the frames in the domain, its combinations with the same noun can be interpreted in multiple ways. The first reason that accounts for this is noun polysemy, illustrated in (1) for dominant systems of English, German and Russian respectively:

(1) a. *The old school is in bad condition.*
         craftsmen of the old school
     b. *Die   alt-e   Schule ist   in schlecht-em Zustand*
         DEF.NOM.F.SG old-NOM.F.SG school be.PRS.3SG in bad-DAT.M.SG state
         'The old school is in bad condition'
         Meister   der   alt-en   Schule
         master.PL DEF.GEN.F.SG old-GEN.F.SG school
         'craftsmen of the old school'
     c. *Staraja   škola   v ploxom   sostojanii*
         old:NOM.F.SG school:NOM.SG in bad:LOC.N.SG state:LOC.SG
         'The old school is in bad condition'
         mastera   staroj   školy
         master:NOM.PL old:GEN.F.SG school:GEN.SG
         'craftsmen of the old school'

In these examples, the word *school* (*Schule* / *škola*) is polysemous and has two different meanings: 'a building where people (usually children) do their studies' and 'a collection of methods for teaching people of a particular profession'. Thus, its combinations with the dominant OLD adjective in each case refer to different concepts: an old (= 'decrepit') building and old (= 'belonging to the past time period') methods of education.

However, noun polysemy is not the only reason for frame ambiguity in noun collocations with the dominant OLD lexeme. The other explanation has to do with the semantic peculiarities of nouns that enable interpretative variations related to the features of anthropocentricity, cyclicity and positive / negative evaluation that we discussed in the previous chapter. For example, according to Rakhilina (1999), the Russian phrase *starye botinki* 'old shoes' is most naturally understood as 'worn out shoes'; however, it may also have a cyclicity-oriented interpretation 'shoes that a person used to wear previously but replaced them with another pair'. The former interpretation matches the second frame ('old clothes'), as it refers to the end of an object's lifespan, while the latter belongs to the third frame ('old boss'), cf. the fully licensed Russian sentence in (2):[3]

(2) *Naden'*   *moi*   *starye*   *botinki,*   *oni*   *ešče*
    put.on:IMP.SG my-ACC.PL old:ACC.PL shoe:ACC.PL they:NOM yet
    sovsem *novye.*
    almost new-NOM.PL
    'Put on my old shoes, they are fairly new.'

The concrete noun *botinki* 'shoes', on the one hand, denotes an object that has a naturally finishing period of effective service; on the other hand, this object may be replaced, depending, for example, on its owner's likes and dislikes, even if it is still in good condition and can be successfully used. Thus, the adjective meaning 'old' can modify both aspects.

Similarly, in collocations where the noun refers to a person, OLD adjectives can be interpreted as referring to age or to other properties, like social status or professional experience.

Interestingly, in such anthropocentric contexts the OLD lexeme modifies these properties differently, at least in Russian and English that we focus on below. Social statuses and positions are perceived as cyclic – occupied by different people at different periods – and fall under frame 3. Professionals, those that are expected to have talent and/or vocation (e.g. teacher, artist, etc.), on the other hand, are normally interpreted in terms of gradual experience gaining. This makes them

---

3.   The author gives credit for this example to Tatiana Bulygina.

compatible with Frame 2, with the only difference that they receive a positive evaluation as not becoming worse in the course of functioning. For instance, the English phrase *old wife* can be interpreted ambiguously: its most natural reading is 'old married woman'. At the same time, the COCA corpus returns the following phrase among the search results to the "old + wife" query:

(3)   *His new wife looks like his daughter. <…> And his old wife is dating a 22-year-old look – alike of him 20 whatever years ago.*

The same type of semantic ambiguity is witnessed in the phrase *old boss*, where the noun can denote both a person and their position: 'aged person in authority' and a 'person who used to occupy an authoritative position' (cf. *former boss*).

Another example demonstrates the interplay of anthropocentricity and positive / negative evaluation, cf. Russian *staryj* (4):

(4)   a.   *Staromu   učitelju   sem'desjat.*
            old:DAT.M.SG teacher:DAT.SG seventy.NOM
            'The old teacher is seventy.'
      b.   *Naš   staryj   učitel'   byl*
            our:NOM.M.SG old:NOM.M.SG teacher:NOM.SG be:PST.M.SG
            strogij,   s   novym   gorazdo legče.
            strict:NOM.M.SG with new:INS.M.SG much   easier
            'Our old teacher was strict, it's much easier with the new one.'
      c.   *Ne mog   staryj   učitel'   otpustit'   ee*
            not can:PST.M.SG old:M.NOM teacher:NOM.SG let_go.INF she:ACC
            bez   naputstvija.
            without   advice-GEN.SG
            'The old teacher could not let her go without advice.'

(4a) and (4b) are interpreted unambiguously, while in (4c) the phrase *staryj učitel'* 'old teacher' can be understood as both 'old in age', 'previous' and also as 'a person who has gained extensive experience.' Thus, the adjective *staryj* 'old' successfully modifies age, status, and experience of the person that the noun *učitel'* 'teacher' refers to. It is also experience that becomes salient in idiomatic phrases like *staryj soldat* 'old soldier' (Russian) or *old soldier* (English).[4]

Depending on noun polysemy of a particular noun, the above listed examples fall into different frames, which are summarized in Figure 2:

---

4.   In both languages, this phrase is used with the reference to a highly experienced person.

Old shoes	Old wife	Old boss	Old teacher	Old school
FRAME 2 ('old clothes')	FRAME 1 ('old person')	FRAME 1 ('old person')	FRAME 1 ('old person')	FRAME 2 ('old clothes')
FRAME 3 ('old boss')	FRAME 3 ('old boss')	FRAME 3 ('old boss')	FRAME 2 ('old clothes', including positive evaluation for professionals)	FRAME 4 ('old coins')
			FRAME 3 ('old boss')	

Figure 2. Polysemy across frames

The fact that the same NPs may have various interpretations serves as an indirect proof of a system dominance and is supported by another evidence of the same kind that comes from the zone of antonyms, where these semantic distinctions are most often lexically marked, as shown in (5):

(5) a. *The six other NFL teams searching for a **new** [as opposed to 'former'] boss had found one.*
     b. *Eaton's arrival has already dented the morale of other Chrysler executives who now see their way to the top blocked by a relatively **young** [in age] boss.*

Thus, for example, in English both *new* and *young* are regarded as antonyms of the dominant lexeme *old* in combination with animate nouns – with *new* referring to cyclic positions and statuses and *young* – referring to age.

5.2   Dominant lexeme and its quasi-synonyms

So far, we have outlined a breakdown of lexical systems into three major classes in order to build an overall typological profile of the OLD domain. It is obvious, however, that when it comes to subtle semantic variations in 'old' + Noun combinations, the combinability rules found in each language appear more complicated. Within the domain of OLD, such complexities can be found and traced in each of the three systems; however, it is the dominant systems which pose a fundamental problem. Logically, the dominant class is expected to be quite straightforward, since its languages feature one lexeme to accommodate all of the frames. However, all languages in our sample demonstrate that there is obvious competition between the dominant lexeme and its quasi-synonyms, and thus the system dominance may be called into question. In this section, we will present examples of statistical analysis that prove helpful in defining distributional correlations between the dominant OLD lexeme and the other attributives in the domain; we will look at English examples and suggest answers to the two questions stated below:

What is the status of the dominant adjective in each frame?

Do the distributional variations within frames influence system dominance?

On the one hand, quantitative evaluations of lexical data allow us to update the questionnaire presented to language experts by supplying additional contexts that are expected to be typologically relevant. On the other hand, quantitative evidence provide extra grounds for classifying a system as either dominant or binary when there is apparent competition between the dominant OLD adjective and its quasi-synonyms. In such cases, each of the alternatives is validated using judgements collected from language experts in the course of dictionary analysis.

In the following, we discuss the example of English, which uses one lexeme for all the four frames and licenses a quasi-synonymous lexeme as a more preferable choice across certain contexts. For example, as was pointed out by the experts, the noun *president,* which exemplifies the 'old boss' frame, is more naturally modified by *former* or *previous* than by the dominant *old* when referring to a shifting political status. Another controversy among the experts when evaluating lexical data belonging to dominant systems was evoked by the context *old wife* presented in the questionnaire: they found it acceptable meaning 'ex-wife' in contrastive phrases like *His old wife was kind-hearted and the new one is beautiful but ill-tempered*, but didn't always agree on whether it could be interpreted as 'ex-wife' in other expressions.

To answer the two questions outlined above, we compared combinability frequencies of the dominant OLD lexeme and its quasi-synonyms with the same set of nouns.

The final list of 9 quasi-synonyms for subsequent statistical analysis comprises the adjectives which are contained in the Oxford English Dictionary [OED] definition of the *old* lexeme and which score high in frequency in the free Google Ngram dataset.[5] We excluded the adjectives that are strongly attracted to one class of objects. Thus, for example, we did not include the adjective *ramshackle*, since it can only be used with nouns denoting buildings.

The resultant quasi-synonyms are listed below (by frames):

'old person'	*aged, elderly, (decrepit)*
'old clothes'	*decrepit, shabby*
'old boss'	*former, previous*
'old coins'	*ancient, antique, archaic*

We harvested the free Google Ngram dataset and elicited bigrams where the first item is an OLD domain adjective and the second one is a noun; we also recorded the raw frequencies of each collocation in the dataset.

---

5. http://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html (accessed on 2019-12-08).

We began by examining the raw counts to see where the competition between the dominant lexeme and its quasi-synonyms is the most pronounced. We established the top frequent Adj + Noun collocations for each quasi-synonym and compared them to the frequencies of the corresponding nouns in combination with the dominant lexeme. The results reveal that the English *old* overwhelmingly dominates the 'old person' and the 'old clothes' frames (see Table 1).

Table 1. *Old* vs. quasi-synonyms: Raw counts for nouns representing Frame 1 and Frame 2

N	Shabby	Old	N	Elderly	Old
clothes	50 272	390 638	people	883 054	2 135 666
house	12 224	1 437 588	man	557 846	18 864 720
N	**decrepit**	**old**	**N**	**aged**	**old**
age	9 726	7 960 504	man	603 683	18 864 720
building	8 236	945 983	woman	367 864	6 299 596

The picture is different for the 'old boss' frame, as shown in Table 2, which compares the frequency distribution of *old* to its quasi-synonyms *former* and *previous*:

Table 2. *Old* vs. quasi-synonyms: Raw counts for nouns representing Frame 3

N	Former	Old	N	Previous	Old
case	1 276 831	59 380	year	3524597	113 976
president	1 056 411	47 069	experience	1003390	42 500
position	339 438	157 224	studies	803836	12 372

The table shows that the *former / previous* + Noun combinations prevail over the *old* + Noun combination in terms of frequency.

A similar result is obtained in the fourth ('old coin') frame (see Table 3). The main competitor of the *old* lexeme within this frame – *ancient* – outscores the dominant lexeme.

Table 3. *Old* vs. quasi-synonyms: Raw counts for nouns representing Frame 4

N	Ancient	Old
times	2 382 368	1 024 236
World	1 216 754	950 428
City	631 370	628 509

The discussion thus far has established that the dominant lexeme is not always preferable compared to its quasi-synonyms. Thus, it would be logical to pose another important question, namely whether the competition between the lexeme and a quasi-synonym in a lexical system may cast doubts on the degree of system

dominance. To glean a clue, we have visualized the distributional patterns of the 10 frequency sets of Ngrams (for the dominant *old* lexeme and its 9 quasi-synonyms) presented above. We took 17 nouns which in combination with *old* and its quasi-synonyms illustrate the four frames of OLD.

The nouns were selected either from the experts' questionnaire or from the examples accompanying the dictionary definitions of the *old* entry [OED].

'old person'	*man, woman, human, person, age*
'old clothes'	*house, clothes, state*
'old boss'	*job, government, wife, friend*
'old coins'	*ritual, tradition, master, paintings, coin*

When viewed out of broader context, these combinations can be interpreted as belonging to more than one frame. However, we observed that with regard to Ngram frequencies each of the nouns shows noticeably higher rates of co-occurrence with the adjective *old* and one or two of its quasi-synonyms; e.g. the noun *government* displays the highest co-occurrence frequencies with the adjectives *old* (430052), *former* (39308) and *previous* (56350), which enables interpreting combinations of *old* with this noun as belonging to Frame 3.

We plotted the natural-logged co-occurrence data on radar charts (Figure 3); each chart represents 17 dimensions corresponding to the 17 nouns which were chosen for visualization (these are marked with labelled straight lines emanating from the center). Thus, we visualize and compare co-occurrence frequencies of a given adjective across all the nouns – the farther from the center the point on each dimension is distanced, the higher the co-occurrence frequency is.
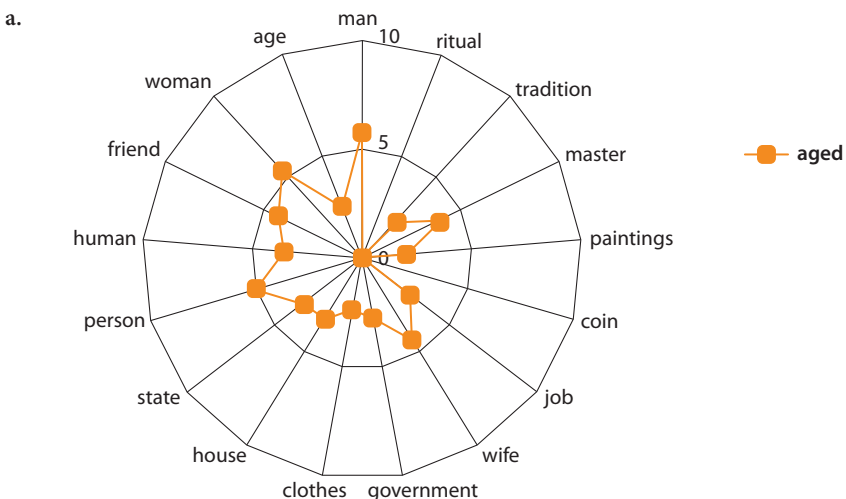
Figure 3a. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams
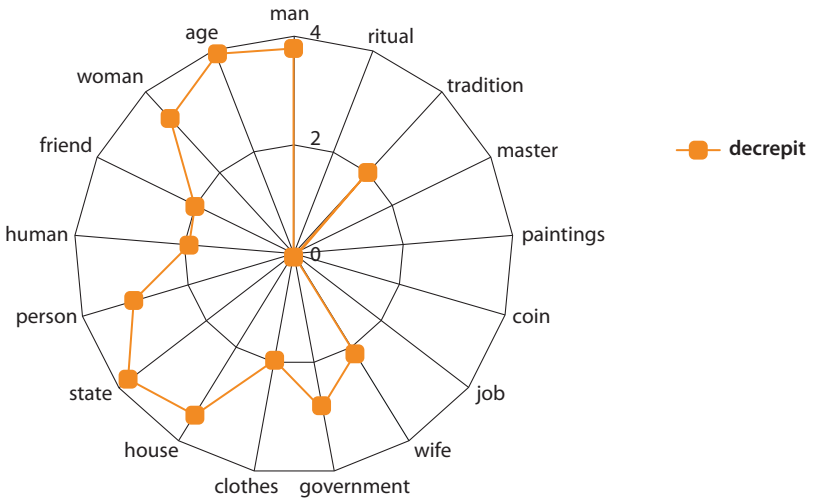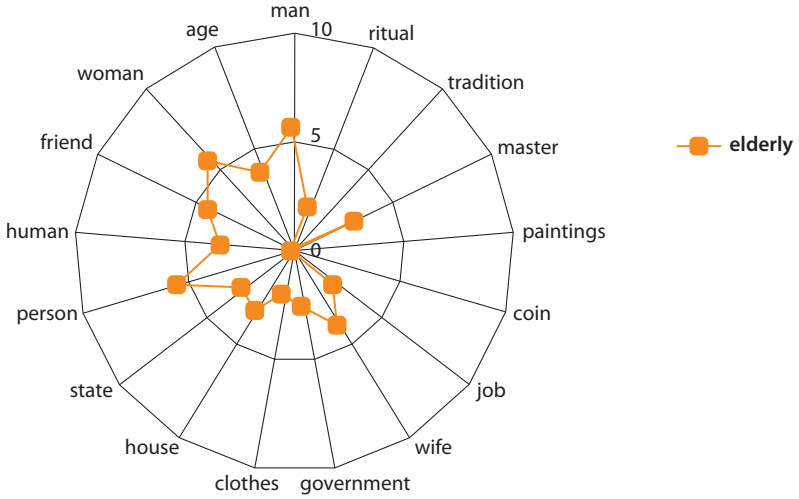
Figure 3b-c. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams

d.
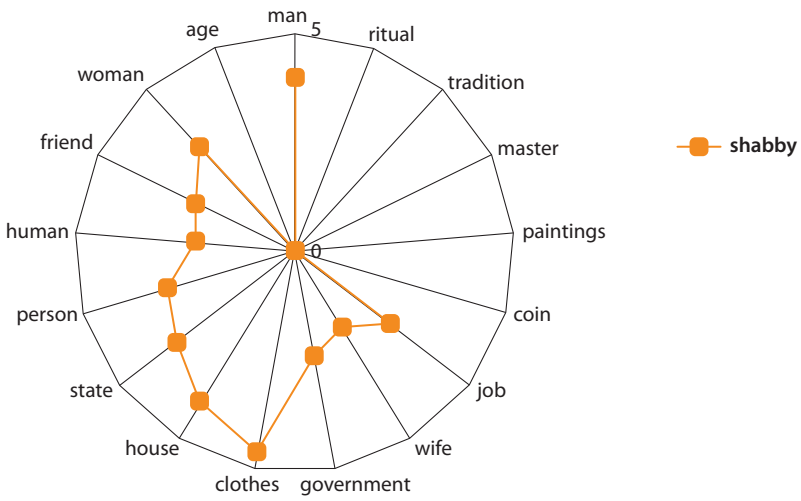
e.
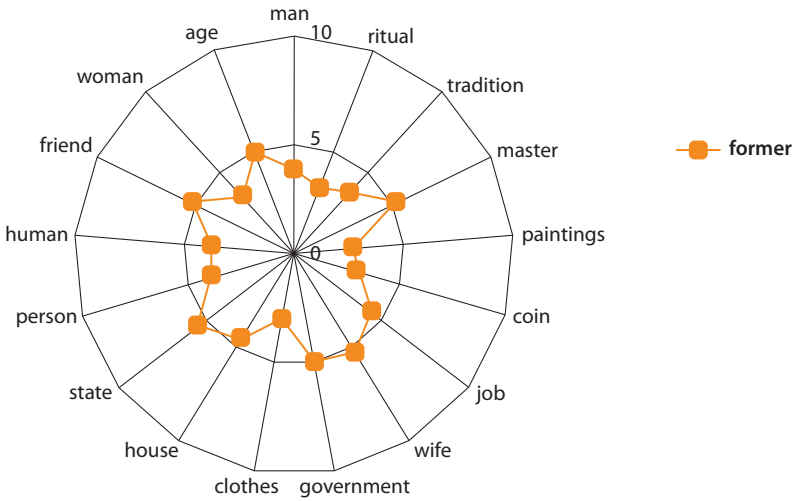
Figure 3d-e. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams
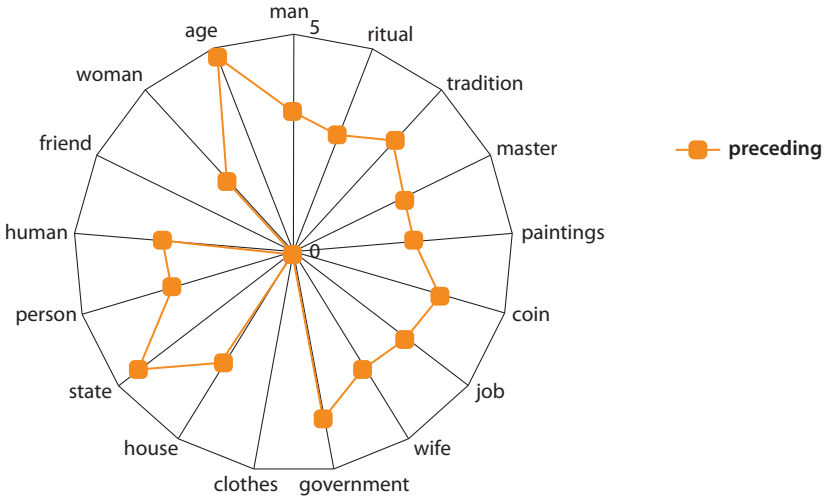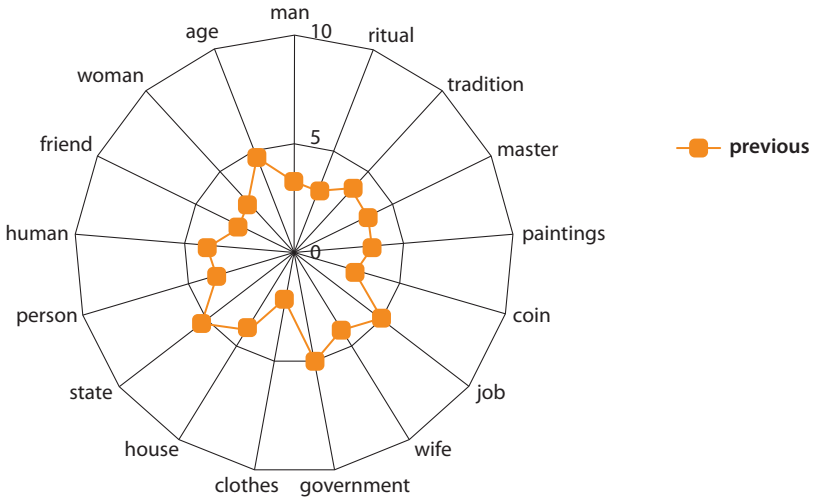
Figure 3f-g. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams

h.
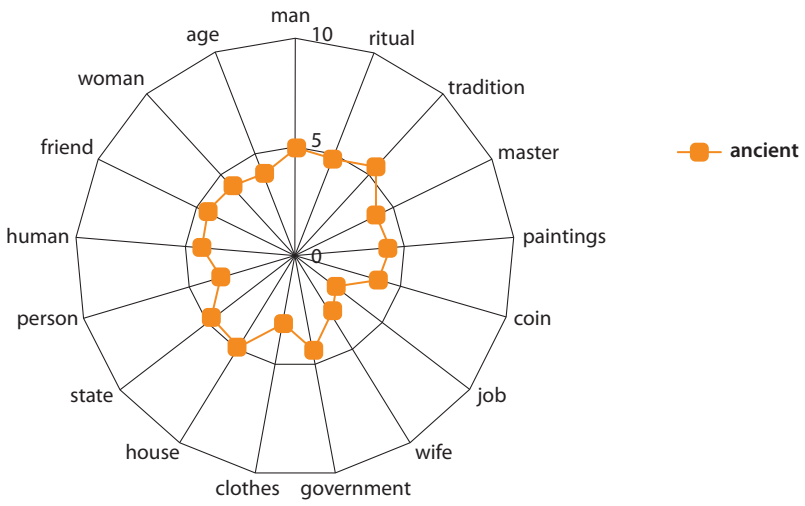
i.
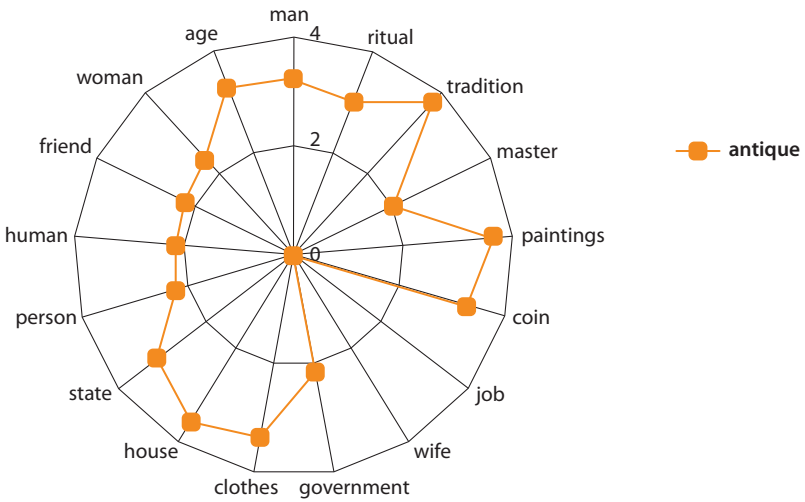
Figure 3h-i. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams

j.

k.
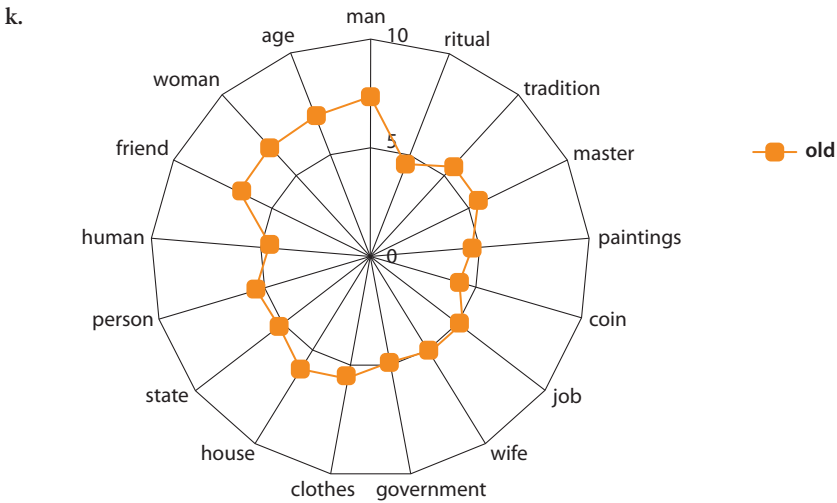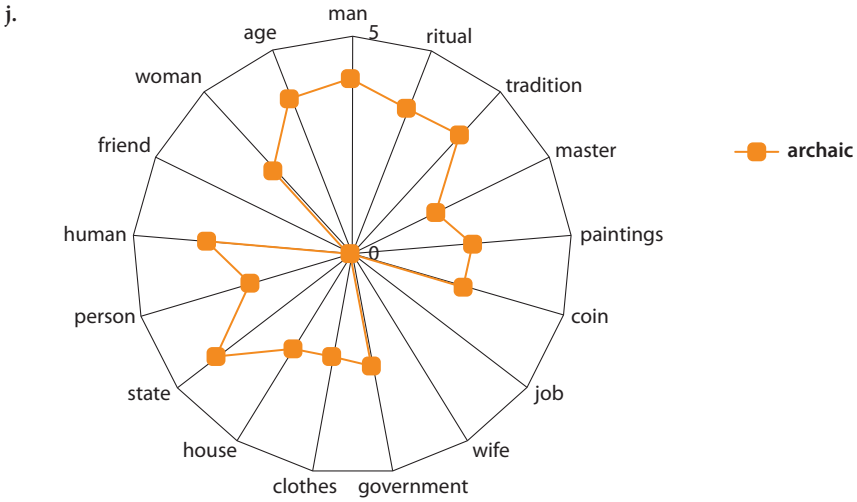
Figure 3j-k. Co-occurrence frequencies for *old* and its quasi-synonyms in noun bigrams

The charts in Figure 3 show how many Adjective + Noun bigrams occur in each da-taset, and offer a general picture of how frames are distributed between *old* and each of its quasi-synonyms. More specifically, the adjectives *aged* and *elderly* combine more frequently with such nouns as *man, woman, person, wife,* and *master* – which proves that they are strongly attracted to the first, 'old person', frame, as opposed to the other three frames ('old clothes', 'old boss', and 'old coins') since very few or no bigrams are attested for *aged* and *elderly* + the nouns *ritual, tradition, painting, coin, job, government, clothes,* and *state.*

Obviously, a more in-depth research is needed to make definitive judgements about the lexical behavior of the adjectives and the preferences of the quasi-synonyms towards each frame. However, the graphs in Figure 3 convincingly illustrate either the uniformity or the lack of uniformity in each adjective's distributional patterns. They also show that despite a certain bias towards the anthropocentric interpretation, *old* is the only adjective that evenly covers all the frames, which yields additional evidence in favor of its dominance in the domain and proves the relevance of the diagnostic contexts included in the questionnaire.

6. Subframes

Until this point, the discussion has revolved around the four frames that constitute the typological profile of OLD. This is, however, a rough picture, and the entire diversity of lexicalization patterns calls for more rigorous inquiry. In this section, we will address this problem through the analysis of lexicalization strategies observed in specific situations.

Apart from frames, there are other colexification patterns that reoccur from language to language and point to typologically relevant strategies. In the present study, we refer to these strategies as subframes. Subframes describe more specific situations than frames do, and thus they are of utmost relevance for creating detailed lists of lexical counterparts across languages. Subframes may obtain special lexical marking or may not receive a specialized means of expression. The latter option is found in the domain of aquamotion, which distinguishes between SWIMMING, SAILING and FLOATING (Maisak & Rakhilina 2007). As for the swimming of birds, this subframe occupies the interim position between the active swimming of people (i.e. SWIMMING frame) and the motion of vessels (SAILING frame), and is extremely rarely accommodated with a special lexeme; rather, this type of swimming is typically expressed with the word denoting either active swimming, or vessel motion. Some instances of specific lexical marking for subframes will be discussed further in this section, as we take a closer look at the structure of OLD.

Within the domain of OLD, certain subframes turn out to be privileged across languages. They either follow a separate lexicalization pattern in binary and distributed systems, or allow of both a dominant and a marginal lexeme in dominant ones. In the latter case, the marginal lexeme is usually regarded as the preferable one.

Before proceeding further, it should be noted that the coverage of subframes for the OLD domain presented below is not meant to be exhaustive; rather the most relevant and recurrent subframes are discussed.

Most of the subframes in the domain are accounted for by the anthropocentricity principle that we discussed in Section 3: out of the four subframes found in the

language sample, three ('oldness grades', 'gender opposition', and 'children's age') are directly related to humans. The fourth subframe indicates the opposition of the concept 'old tree' (labelled here as 'old oak') to the concept 'old man'.

6.1   'Old person': Oldness as a gradual property

Oldness is a cumulative, gradually acquired property, and this intrinsic characteristic manifests itself in lexical oppositions. In a number of languages, separate lexemes accommodate different degrees of oldness, as in Figure 4:

Not very old < Old < Very old

Figure 4.   Lexicalization of oldness degrees

Russian, for example, distinguishes between old people (*staryj čelovek* 'old person'), very old and likely to be disabled ones (*prestarelyj* < *staryj* 'old'), and people who are advanced in years (*požiloj* 'aged'). Interestingly, the latter term is often used to speak politely about seniors and to avoid connotations with physical or mental disability that are entailed by the dominant lexeme *staryj* 'old'.

6.2   'Old person': Gender opposition

Systems where gender oppositions are marked either lexically or grammatically are frequently found across languages (cf. Corbett 2013 and Hellinger & Bussmann 2001, among others); the OLD domain is not an exception.

Special lexemes for referring to men's and women's age are found, for example, in the binary system of the Maidu language (Maiduan, North America), where the word *wajsi* is used to talk about men and the word *kylokbe* about women. This opposition also holds for verbs, cf. *nenodoj* ('to get aged') and *lokbekbedoj* 'to get aged (of women)' (see Dixon 1912). Other languages that have a special lexeme to refer to women include Nenets (Uralic, distributive system), Kurdish (Iranian, distributive), or Tajik (Iranian, binary).

Gender oppositions may also affect lexical choice in polite contexts mentioned in Section 6.1. In Russian the word *požiloj* 'aged' is equally frequent when referring to men and women;[6] however, in the contexts describing men, *staryj* is much

---

6.   Raw frequencies from the Russian National Corpus for *požiloj* used with nouns denoting men total 7119 uses, and with nouns denoting women – 7342 uses. The same opposition for *staryj* number 49 (men) as compared to 632 (women) (available at http://ruscorpora.ru, accessed on 2021-12-17).

less preferable regardless of whether the man is distinctly old or less advanced in years.[7]

People's age is described by the first frame; it is expressed by the dominant lexeme in Russian (*staryj čelovek* 'old man', *starye ljudi* 'old people', *staryj professor* 'old professor', *bednaja staraja ženščina* 'poor old woman'). However, in combination with the word *mužčina* ('man', masculine gender marked) it is not acceptable and should be substituted by a quasi-synonym, cf. (6):

(6) *Očarivateľnyj   požiloj   mužčina   privëz   eë*
    charming:NOM.M.SG aged:NOM.M.SG man:NOM.SG deliver:PST.M.SG she:ACC
    v 16.30   (RNC)[8]
    at 16.30
    lit. 'A charming aged man delivered her at 16.30'.

6.3 'Old person': Speaking of children

Another situation that lexical systems are sensitive to when it comes to the OLD domain is the description of the age of children.

In relation to this type of referents, it is important to distinguish between absolute and comparative age. In terms of absolute age, children stop being children when they reach a maturation period, which starts long before they enter their oldness. Thus, they do not get old in a commonly accepted sense and logically the combination of two words – 'old' and 'child' is impossible. None of the languages in the sample allowed for combinations of the word meaning 'child' with the word that measures his/her age against the oldness scale.

Within the 'children' subframe, we also consider nouns for family members that hold a strong association with the childhood period. Such nouns include sibling names (*sister / brother* = 'those who one normally knows from childhood', *son / daughter / nephew / niece* = 'one's children or children of one's sibling'). None of the systems exemplified in our sample exhibited the possibility of coding absolute age of such referents with the adjective meaning 'old': * *old brother*, * *old son*, * *old nephew* (referring to an aged relative).

However, in a limited number of languages, it is possible to use such adjectives for scaling comparative and auto-comparative age.

Using OLD domain lexemes for comparing siblings' age – referred to here as comparative age – is attested in, for example, Mandarin Chinese (binary system),

---

7. This restriction, however, does not hold for negative contexts, like *staryj* 'old' vs. *ešče ne staryj* 'not old yet'.

8. Russian National Corpus (available at http://ruscorpora.ru, accessed on 2018-11-15).

where the adjective 老 *lǎo* 'old' forming a disyllabic compound in combination with parametric words and ordinals is used for ranging siblings against the age scale:

(7) 老 *lǎo* 'old' + 大 *dà* 'big' = 'the eldest sibling'
     老 *lǎo* 'old' + 二 *èr* 'second' = 'the second eldest sibling'   (referring to age)

A similar strategy is relevant for English. For a more advanced in age sibling English employs a comparative degree of the adjective *old* – a dominant lexeme of the domain:

(8) *my elder/eldest brother*
     my elder/eldest son
     my elder/eldest nephew

Auto-comparative age – speaking about the age of same referent in different points of time – is a situation that is normally expressed with a word from a different domain (cf. *grown up* in English or *vz=ros=lyj*[9] in Russian). Nonetheless, that does not mean that using the OLD lexeme in this case is not possible.[10] A case in point is Hawaiian, which lexicalizes the idea of a child / person proceeding in age with the help of the *old* adjective.

(9) *ʻemakule* (v.*)* – to decay, to grow old
     ʻelemakule (adj.) – normally meaning 'decrepit'
     he hopena ʻelemakule 'the result of being an old man' <…> said jokingly of
     oneself as he advances in years   (Elbert, Pukui 1986)

6.4   'Old person' vs. 'old oak'

The last subframe to be discussed is labelled here as 'old oak'. According to the sample data, trees are sometimes granted a special lexical status in the domain, otherwise they are colexified with people, i.e. described by the lexeme used for naming a person's age.

The following lexicalization patterns are found in binary and distributive systems:

–   Turkmen, binary system: *garry* (of a man and an oak);
–   Ossetian, distributed system: *zærond* (old man, old woman, old oak);
–   Georgian, distributed system: *beberi* ('old' for plants, pejorative of a person).

---

9. The Russian prefix *vz-* generally expresses the meaning 'up', and the root *-ros-* means 'grow'.

10. Note that in English this idea is expressed with a construction *to get older*, part of which is the dominant lexeme.

Thus, trees may be conceptualized either as living beings or inanimate objects and illustrate another specific situation that is expressed with recurrent lexicalization patterns.

7.   Conclusion

Introducing more languages into a sample for cross-linguistic research of the lexicon has long been a strong wish of lexical typologists. In this research, we made an attempt to expand an initial sample of 8–10 languages commonly scrutinized in frame-based studies of lexical systems, and conducted an extensive dictionary study of OLD terms in 78 languages.

The dictionary data were subjected to additional expertise: the collected instances were analyzed, classified, and forwarded to language experts for ultimate assessment.

The results demonstrate that thus collected larger data samples yield relevant typological classifications for well-structured and previously described semantic domains. They enable finding both recurrent colexification patterns and limitations in colexification strategies.

Dictionary analysis and the expert assessment provided grounds for classifying lexical systems of the OLD domain into three types: dominant (one lexeme covering all the four frames), binary (two lexemes covering all the four frames), and distributed (more than two lexemes covering all the four frames).

Special attention has been paid to the structure of dominant systems and the status of the dominant lexemes. The notion of dominance was analyzed with regard to English using counts of bigram distribution with the dominant lexeme of OLD and its quasi-synonyms.

The paper also gave an account of cross-linguistically recurrent colexification patterns – the so-called subframes – and ambiguous readings of 'old' + Noun phrases that fall into different frames. All these taken together present a holistic picture of the OLD domain from a typological perspective.

Acknowledgements

Abbreviations

3	the 3rd person	LOC	locative
ACC	accusative	M	masculine
DAT	dative	N	neutral
DEF	definite	NOM	nominative
F	feminine	PL	plural
GEN	genitive	PRS	present
IMP	imperative	PST	past
INF	infinitive	SG	singular
INS	instrumental		

References

Beard, R. 1991. Decompositional composition: the semantics of scope ambiguities and 'bracketing paradoxes'. *Natural Language and Linguistic Theory* 9: 195–229.
https://doi.org/10.1007/BF00134676

Corbett, G. 2013. Sex-based and Non-sex-based Gender Systems. In *The World Atlas of Language Structures Online*, M. S. Dryer & M. Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology, <http://wals.info/chapter/31> (6 April 2020).

Croft, W. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

Dahl, Ö. 1970. On sets and prepositional functions in grammar. *Scando-Slavica* 16: 75–96.
https://doi.org/10.1080/00806767008600543

Dixon, R. B. 1912. *Maidu texts. Publications of the American Ethnological Society* 4. Leiden: Brill Publishers.

Elbert, S. H. & Pukui, M. K. 1986. *Hawaiian Grammar*. Honolulu: University of Hawaii Press.

Georgakopoulos, T. & Polis, S. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2): 1–33.
https://doi.org/10.1111/lnc3.12270

Haspelmath, M. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In *The New Psychology of Language*, M. Tomasello (ed), 211–242. Mahwah, NJ: Erlbaum.

Hellinger, M. & Bussmann, H. (eds). 2001. *Gender Across Languages*. Amsterdam: John Benjamins.
https://doi.org/10.1075/impact.9

Maisak, T. & Rakhilina, E. (eds). 2007. *Glagoly dviženija v vode: leksičeskaja tipologija*. [Verbs of Aqua-Motion: Lexical Typology]. Moskva: Indrik.

[OED]: Simpson, J. A. & Weiner, E. S. C. (eds). 1989. *The Oxford English Dictionary*. Oxford: Oxford University Press.

Rakhilina, E. 1999. Aspectual classification of nouns: A case study of Russian. In *Tense-aspect, transitivity and causativity*, A. Werner & L. Kulikov (eds), 341–350. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.50.23rak

Rakhilina, E. & Reznikova, T. 2016. A Frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–130. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110377675-004

Rakhilina, E. V., Ryzhova, D. A. & Badryzlova, Yu. G. In press. Lexical typology and semantic maps: perspectives and challenges. *Zeitschrift für Sprachwissenschaft*.

Taylor, J. R. 1992. Old problems: adjectives in cognitive grammar. *Cognitive Linguistics* 3: 1–35. https://doi.org/10.1515/cogl.1992.3.1.1

Appendix

Questionnaire used for expert assessment

1. [Old] people get cold easily.
2. [Old] women look younger than [old] men.
3. His [old] wife was kind-hearted, and the new one is beautiful but bad-tempered.
4. Remarkable works of art by [Old] Masters are kept in museums.
5. There is an [old] oak close to our house; it is probably 200 years old.
6. We have an [old] raincoat in the closet, it's time to throw it away because it leaks.
7. An [old] friend will always lend a hand.
8. Our [old] boss let us finish the working day early, but the new one never does so.
9. The [old] coins are made of silver.
10. [Old] computers were very slow.
11. [Old] fridges were very noisy.
12. [Old] wine is expensive because it has been aged for a long time.
13. We still have lots of [old] potatoes left from the previous year.
14. [Old] memories always make you feel sad.
15. [Old] times are all gone, and people are different now.
16. Now I live in a new apartment, and no one lives in the [old] one; the house will soon be demolished.

Link to the online map with the typological results and the full list of dictionaries used: http://lcph.bashedu.ru/cgi-bin/oldmap.html

Talking temperature with close relatives

Semantic systems across Slavic languages

Maria Koptjevskaja-Tamm
Stockholm University

The chapter compares the temperature adjectives ('hot', 'cold' etc.) across Slavic against a broader typological background. The comparison targets both the *systems as a whole* and the *forms* involved in them. The main questions are how (dis)similar the temperature systems of closely related languages can be, and what is stable vs. changeable in the temperature terms of closely related languages. Slavic languages show substantial cross-linguistic variation in their systems (ranging from two to four main temperature values), while on the whole confirming several earlier tentative generalizations in Koptjevskaja-Tamm (2015). The temperature terms themselves differ in stability, both in meaning and in form (with 'warm' being the most stable term on both counts), even though most of them are traceable to proto-Slavic and even to proto-Indo-European.

Keywords: temperature, Slavic, lexical stability, semantic change, semantic system predicative

1.   Introduction

Typically, closely related languages share many properties with each other, but also diverge in various respects. There is a long tradition of intra-genetic studies examining synchronic and diachronic lexical divergence, especially dealing with semantic change and lexical replacement (see Koch 2016 for an illuminating survey). As noted in Majid & Dunn (2015: 2), these studies normally focus on individual words, rather than on whole lexical fields or semantic domains. The last decade's rapid developments in semantic and lexical typology have made it possible to compare lexicalization of whole semantic domains in more systematic and cross-linguistically encompassing ways. While the main interest of the discipline probably lies in comparison of unrelated languages, several recent studies have compared lexicalisation of semantic domains among closely related languages. These have included study

of expressions for LOCATION in the two Mayan languages Tzeltal and Yucatec Maya (Bohnemeyer & Brown 2007), verbs of ROTATION in Russian and Polish (Rakhilina 2010), expressions for the CUT and BREAK domain across Germanic (Majid et al. 2007), and AQUAMOTION verbs across Germanic and Slavic (Koptjevskaja-Tamm et al. 2011). All of them demonstrate various degrees of lexico-typological differences between lexical systems in closely related languages.

The fact that even closely related languages may display significant divergence in their lexical systems does not come as a surprise. To quote Rakhilina & Reznikova (2016: 101), "[w]hile grammatical constructions take centuries to evolve, vocabulary is much more fluid. A single generation of speakers may witness words falling in and out of use and word meanings changing dramatically. As a result, even such close relatives as Russian and Polish do not necessarily have many cognates in a given domain, and even when they do, such words tend to have meanings quite dissimilar to those of their "cousins"". Accordingly, careful comparison of the lexical systems in closely related languages can be a window onto possible ways in which lexical systems may arise and develop.

In line with these developments in lexical typology, the present chapter compares lexicalization of the semantic domain of TEMPERATURE in the Slavic languages. It builds on the already extensive lexico-typological research on the linguistics of temperature (cf. Koptjevskaja-Tamm (ed) 2015), encompassing more than 50 genetically, areally and structurally diverse languages. This research has unveiled the amazing cross-linguistic diversity in how languages carve up the temperature domain by means of their linguistic expressions, but also the systematicity behind it. The chapter focuses on central temperature adjectives, such as those for 'hot' or 'cold' etc., attending to the similarities/dissimilarities in the Slavic semantic systems as seen against a broader typological background. In this undertaking, the chapter elaborates on the diachronic aspect of cross-linguistic comparison to a larger extent than what has been done in the studies mentioned above. Notably, the comparison targets both the *systems as a whole* and the *forms* involved in them, and the overarching issue here is *intra-genetic constancy* vs. *variation*, or *stability* vs. *change*. Or, to quote Frans Plank (2010), what about the lexicon and grammar of temperature expressions is stable and unstable, and how and why the temperature meaning can change. The more specific research questions include the following:

i.   How (dis)similar can the temperature systems of closely related languages be?
ii.   What is stable vs. changeable in the temperature terms of closely related languages?
     a.   From where do temperature expressions come?
     b.   To what extent do genetically related languages share temperature cognates?

    c.   If they do, do the cognates have the same or similar meanings?

    d.   How can the meaning of temperature expressions change within
        the temperature domain proper?

    e.   How are temperature expressions replaced (by other native expressions
        or by borrowings) and what happens with the older expressions?

iii. What is the role of language contact in shaping the temperature-term systems?

The chapter is based on data from multiple sources – earlier descriptions of the temperature systems in Russian (Rakhilina 2000; Koptjevskaja-Tamm & Rakhilina 2006), Ukrainian (Kryvenko 2015) and Serbian (Rasulic 2015), data collected according to the temperature guidelines in www.ling.se/staff/tamm/tempquest.pdf and subsequent discussions with language experts and native speakers, searches in monolingual and parallel corpora, and synchronic and diachronic dictionaries. As will occasionally become clear from the text, sources differ considerably in their reliability and commensurability with each other. Synchronic dictionaries are generally not sufficient for capturing the nuances in the meanings and uses of the temperature terms that are central for this study, in particular, when it comes to drawing the borderlines between terms with fairly similar meanings. Etymological dictionaries can vary in their reconstructions, in particular when it comes to the meaning of the suggested source. Corpora vary in their size and genre composition, etc. The investigation has therefore required careful triangulation among the different methods, and discussions with language experts and native speakers about the interpretation and the exposition have played a major role.

    The next section introduces general background on the cross-linguistic comparison of temperature systems ("the linguistics of temperature"). Section 3 is devoted to a relatively detailed description of the Russian system of temperature adjectives, which serves as the point of departure for the following comparison across Slavic. Sections 4–6 turn to the Slavic adjectives for the warming zone, Sections 7–8 revolve around those for the cooling zone, and Section 9 finally addresses the in-between temperatures ('lukewarm'). Section 10 summarizes the main findings with respect to intra-genetic constancy vs. variation, or stability vs. change formulated above. Finally, Section 11 puts these findings in a broader cross-linguistic context.

2.   Introducing the linguistics of temperature

Languages have different kinds of expressions with temperature meanings that can be used for various functions. Of primary interest for our chapter are expressions that denote temperature states or properties and are used for temperature predication, such as *The night is COLD*, and for temperature modification, such as *You will need a WARM blanket for COLD nights*.[1]

As confirmed by the chapters in (Koptjevskaja-Tamm (ed) 2015), cross-linguistic variation in categorization of the temperature domain can be captured by three main dimensions: TEMPERATURE VALUE, FRAME OF TEMPERATURE EVALUATION, and EVALUATED ENTITIES.

The most obvious dimension is TEMPERATURE VALUE, which covers the distinction between warming and cooling temperatures, the distinction between excessive heat and pleasant warmth, etc. Several different scales have been suggested for evaluating human temperature sensation for various purposes – e.g., the Bedford 'comfort scale' (Bedford 1936) and the ASHRAE (The American Society of Heating, Refrigerating and Air-Conditioning Engineers) 'sensation scale' (http://www.ashrae.org/), used for correlating thermal sensation with human thermal environmental conditions, or Herbert Hensel's scale (Hensel 1981), designed for measuring physiological and psychological responses to entities with different temperature qualities (cf. Koptjevskaja-Tamm 2015: 12 for details). In an attempt to generalize the descriptions for the different values adopted in these scales, as well as typical explications and translations of temperature words in sources on different languages, I will be using the following semantic labels as more or less covering the meanings given in parenthesis:

> 'cold' ('generally / unpleasantly / uncomfortably cooling'),
> 'cool' ('moderately / pleasantly / comfortably cooling'),
> 'warm' ('pleasantly / comfortably warming'),
> 'hot' ('unpleasantly / uncomfortably / dangerously warming').

The dimension FRAMES OF TEMPERATURE EVALUATION covers the distinctions among tactile temperature (*The stones are cold*), ambient temperature (*It is cold here*), and personal-feeling temperature (*He feels cold*) (cf. Plank 2010), with a possible addition of illness or bodily-symptoms temperature (*He has fever*).[2]

---

1.   As a result, we exclude other kinds of temperature expressions such as expressives, ideophones, and interjections used for depiction (e.g. *brrr* for 'cold'), expressions for processes and actions whereby an entity acquires a certain temperature or an agent changes its temperature (e.g., *The house has COOLED OFF by a few degrees* or *We CHILLED the drinks on ice*), and expressions for temperature reference (e.g. *We suffered from the COLD*).

2.   Extended uses of temperature terms, e.g., for describing emotions and human propensities, like *a warm person* in English, are not considered here.

Finally, some EVALUATED ENTITIES, for instance water, may require particularly elaborated subsystems of temperature expressions with additional terms for extreme temperature values, such as 'ice-cold' and 'boiling hot', or with a special term for in-between temperatures such as 'lukewarm'.

A striking feature of temperature systems across languages is their internal heterogeneity in that their different parts may behave differently in lexical choice and/or morphosyntactic patterns or differ in the number of temperature-value oppositions. As suggested in Koptjevskaja-Tamm (2015: 19), personal-feeling temperature terms often show a reduced set of temperature-value oppositions, or a lower degree of elaboration, as compared to terms for tactile temperatures. In turn, tactile temperature terms may be less elaborated than ambient temperature terms. The hierarchy in Figure 1 depicts these relationships, where ≤ indicates the relative number of lexical terms in the subdomain and the arrow head represents the highest number of lexical terms.

PERSONAL-FEELING ≤ TACTILE ≤ AMBIENT

number of temperature-value oppositions
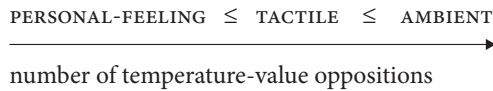
Figure 1. The hierarchy of temperature-value elaboration within the temperature subsystems

However, even when the lexical subsystems used for temperature evaluation and/ or for different entities differ in their internal elaboration, we may still talk about the *main temperature-value distinctions* for the whole system, referring to the distinctions that are shared by all the frames of temperature evaluation (but not necessarily expressed by the same means, as explained in the next paragraph). The most common temperature-value systems represented in Koptjevskaja-Tamm's (2015) volume are summarised in Figure 2, where the notation 'x'/'y' means that the language uses the same term for the temperature values corresponding to 'x' and 'y', i.e. to the whole warming and to the whole cooling zone.[3]

two-value systems	'cold'/'cool'	'warm'/'hot'	
three-value systems	'cold'/'cool'	'warm'	'hot'
four-value systems	'cold' 'cool'	'warm'	'hot'

Figure 2. The main temperature-value systems represented in the languages in Koptjevskaja-Tamm (ed) (2015)

---

3. An additional and relatively infrequent one-value system, according to Ameka (2015), is found in the Kwa language Ewe. Ewe does not have a dedicated temperature term (or 'basic term' in Ameka's terminology) for 'warm'/'hot'.

The relatively restricted typology presented in Figure 2 does not account for all cross-linguistic variation. At this point it is reasonable to introduce the distinction between *domain-central* and *domain-peripheral* terms. *Domain-peripheral* terms are typically terms for extremely hot or extremely cold temperatures (e.g., *scorching, ice-cold*), or for intermediate temperatures (e.g., *tepid*), that are restricted to a few particular entities (for example sun, fire, water). Here languages vary quite a lot. *Domain-central* terms, on the other hand, have a broader application, covering either a particular frame of temperature evaluation (*frame-specific* terms), or pertaining to several frames (*cross-frame* terms) or even to all frames (*frame-neutral* terms). For instance, 'cold' in Russian is referred to by the frame-neutral domain-central adjective *xolodnyj*, while 'hot' is expressed by two domain-central adjectives: the frame-specific term *gorjačij* for tactile temperatures and the cross-frame term *žarkij* for non-tactile temperatures – i.e., for ambient and personal-feeling temperatures. Languages vary quite a lot as to whether they have frame-specific vs. cross-frame vs. frame-neutral terms to express different temperature values.

Moreover, the semantic definitions of 'warm' vs. 'hot' or 'cool' vs. 'cold' given in the first part of this section are too rough and approximate, and languages may vary as to where the borderlines between the two members in each lexical pair are actually drawn.

Finally, there is the issue of the morphosyntactic properties of different temperature expressions, such as their part-of-speech affiliation and the constructions they are used in.

The next section will detail the system of the central temperature adjectives in Russian. This will later serve as the point of departure for discussing the intra-genetic comparison of the temperature adjective systems across Slavic.

3.   The temperature adjectives in Russian

Table 1 shows the domain-central temperature adjectives in Russian, representing a version of the four-value system sketched in Figure 2. The domain-central adjectives may be used both in attribution and in predication, with interesting morphosyntactic properties corresponding to the different frames of temperature evaluation (cf. Rakhilina 2000 and Koptjevskaja-Tamm & Rakhilina 2006 for additional details on the semantics of the Russian temperature terms in attribution).

Table 1. The domain-central temperature adjectives in Russian

'cold'	'cool'	'warm'	'hot'	
			'tactile'	'ambient', 'personal-feeling'
xolodnyj	proxladnyj	teplyj	gorjačij	žarkij

For tactile evaluation, the temperature adjectives follow the standard (canonical) morphosyntactic patterns of qualitative adjectives: they easily appear in adnominal attribution (1a) and are frequently used as adjectival predicates, agreeing with the subject in number and gender and requiring a copula in the non-present tense (1b).

Russian
> (1) Tactile temperature: 'cold'
>> a. Attribution
>> *xolodn-aja* *kastrjulj-a*
>> cold-NOM.F.SG pan-NOM.SG
>> 'a cold pan'
>> b. Predication
>> *Kastrjulj-a* *byl-a* *xolodn-aja/oj.*
>> pan-NOM.SG was-F.SG cold-NOM.F.SG/INS.F.SG
>> 'The pan was cold.'

Things are more complicated for ambient and personal-feeling evaluation. For reference to ambient temperatures, temperature adjectives can collocate with expressions denoting weather and climate, periods of time, geographic spaces and in-door spaces, sources of heat, etc., both as adnominal attributes (2a) and as canonical adjectival predicates (2b). Ambient temperatures are in these cases construed as properties of particular entities, however abstract those may be. Predication of ambient temperatures is also often – and more generally – expressed by an impersonal construction involving the form called "predicative" in the Slavic linguistic tradition (Ru *kategorija sostojanija*), as in (2c).[4] I will use the terms "quasi-referential" vs. "non-referential ambient temperature" for the contrast between (2b) and (2c).

---

4. The status of predicatives as inflectional vs. derivational forms and, consequently, their word-class status is disputed. In this study I have chosen to ignore this problem by treating the temperature predicative forms as predicative alternatives of the temperature adjectives.

(2) Ambient temperature: 'cold'
- a. Attribution
  xolodn-aja   zim-a / pogod-a
  cold-NOM.F.SG winter-NOM.SG / weather-NOM.SG
  'a cold winter, weather'
- b. Predication (quasi-referential)
  Zim-a/pogod-a   byl-a   xolodn-aja/oj.
  winter-NOM.SG / weather-NOM.SG was-F.SG cold-NOM.F.SG/INS.F.SG
  'The winter / weather was cold.'
- c. Predication (non-referential)
  Tam / Včera / Zimoj   byl-o   xolodn-o.
  there / yesterday /at_winter was-N.SG cold-PRED
  'It was cold there / yesterday / at winter.'

I have chosen the term "quasi-referential" for (2b), rather than "referential", in order to underline the fact that predication is often made about very abstract entities, such as weather and climate, or time periods, which are thus construed as having a particular temperature. In (2b), the predication is about winter, which has the property of being cold, whereas in (2c), 'at winter' is a specification of the circumstances under which it was cold.

Personal-feeling temperatures are primarily expressed by impersonal predicative constructions, akin to those for non-referential ambient temperatures, but containing an additional experiencer phrase in the Dative case (3). Temperature adjectives are not completely excluded as attributes to nominals denoting persons, but in these cases, they are used metaphorically (literally 'a cold person' indicates 'emotionally not responsive') or, much more rarely, they are used for reference to tactile evaluation (literally 'a cold baby' indicates 'a baby who feels cold by touch').

(3) Personal-feeling temperature: 'cold'
  Mne /Pet-e   byl-o   xolodn-o.
  I:DAT /Peter-DAT was-N.SG cold-PRED
  'I / Peter felt cold.'

The five different constructions (Cx) involved in temperature predication in Russian are summarized in Figure 3. The meanings (or definitions) are inspired by the Fillmorian FrameNet explications (https://framenet.icsi.berkeley.edu/) and, broader, by Construction grammar. The headed arrows in the diagrams refer to inheritance links among constructions, in this case, instance links, whereby a more schematic item in one construction is specified by a more concrete (or substantive) item in the construction connected to it by a headed arrow (cf. Goldberg 1995: 75–81; Koch 2012: 559–560).

Standard temperature property predication: Cx₁	Meaning:	An Entity has a particular temperature
	Form:	NP_NOM (COP_AGR*) TempAdj_NOM/INSTR**·AGR*

Tactile temperature predication: Cx₁.₁	Meaning: tactile temperature Form: NP is concrete and does not refer to "ambient" entities; TempAdj can refer to tactile temperatures	Quasi-referential ambient predication: Cx₁.₂	Meaning: ambient temperature Form: NP refers to "ambient" entities; TempAdj can refer to ambient temperatures

Non-referential ambient temperature predication: Cx₂	Meaning:	The temperature in a certain environment, determined by time and place, is specified.
	Form:	(Place) (Time) (COP_3SG/N.SG) TempPred TempPred can refer to ambient temperatures

Personal-feeling temperature predication: Cx₃	Meaning:	An experiencer senses different degrees of warmth that may or may not be related to the ambient temperature.
	Form:	NP_DAT (COP_3SG/N.SG) TempPred TempPred can refer to personal-feeling temperatures.

Figure 3. Constructions for temperature predication in Russian

Notes:

*AGR = the copula verb (absent in the Present) and the predicative adjective agree with the subject in number and person or gender (depending on the Tense)

**NOM/INS = when there is no copula, i.e., in the Present, predicative adjectives take the Nominative case; in other contexts, they appear either in the Nominative or in the Instrumental case, depending on a complex interaction of various factors.

The adjectives *xolodnyj* 'cold', *proxladnyj* 'cool' and *teplyj* 'warm' can be used in all the contexts and morphosyntactic constructions described above and are therefore frame-neutral. For 'hot', however, Russian distinguishes between two adjectives – the tactile (frame-specific) adjective *gorjačij* seen in (4a) and the ambient/personal-feeling, i.e. non-tactile (cross-frame) adjective *žarkij* seen in (4b–d).

(4) 'Hot' as predicate

    a. Tactile predication

       Kastrjulj-a   byl-a   gorjač-aja/ej.

       pan-NOM.SG was-F.SG hot_tact-NOM.F.SG/INS.F.SG

       'The pan was hot.'

    b. Quasi-referential ambient predication

       Vesn-a/Pogod-a   byl-a

       spring-NOM.SG /weather-NOM.SG was-F.SG

       žark-aja/oj.

       hot_nontact-NOM.F.SG/INS.F.SG

       'The spring/ weather was hot.'

    c.   Non-referential ambient predication

       Tam / Včera   *byl-o*   *žark-o.*

       there / yesterday was-N.SG hot_nontact-PRED

       'It was hot there / yesterday.'

    d.   Personal-feeling predication

       Mne /Pet-e   *byl-o*   *žark-o.*

       I:DAT /peter-DAT was-N.SG hot_nontact-PRED

       'I / Peter felt hot.'

To sum up, the Russian temperature-term system is sensitive to distinctions among the frames of temperature evaluation, both in the morphosyntactic patterns and in the lexical choices within the 'hot' zone.[5] We will return to this topic in Section 5.

However, since all the frames of temperature evaluation manifest the same distinctions among 'cold', 'cool', 'warm' and 'hot' (see Table 1) we can still argue that the Russian temperature terms are organized in a four-value system.

Elaborating on the temperature-value distinctions themselves, the main temperature contrast in Russian is between 'cold' and 'hot', seen in conventionalized pairs of expressions, such as those involving dishes, drinks, water taps, etc. As further argued in Koptjevskaja-Tamm & Rakhilina (2006), *teplyj* 'warm' describes warming temperatures within a relatively narrow temperature zone, while both *gorjačij* and *žarkij* cover a wide range of temperatures, from "pleasant" warming to any unpleasant, dangerous burning ones, with no upper limit. The reference of *teplyj* can hardly be raised by intensifiers, such as 'very' or 'extremely'. The temperature in *očen' teplaja komnata* 'a very warm room' is not significantly higher than that in *teplaja komnata*, and *očen' teplye ruki* 'very warm hands' does not imply that the person has fever. The denotation is more or less the same with and without *očen'*, but there is a pragmatic difference: what is intensified is the evaluative component of cosiness, comfort and pleasantness. The border between *teplyj* and the 'hot' adjectives is therefore quite fixed and obligatory.[6] We will shortly elaborate on the status of *proxladnyj* 'cool'.

---

5.   Of the two 'hot' adjectives, only the non-tactile one has a corresponding noun, *žara* (also *žar* in some of its uses), denoting high ambient temperatures. This is in line with the other three domain-central temperature adjectives, the nominal equivalents of which (*teplo, proxlada, xolod*) are primarily used to denote ambient temperatures. I am grateful to Daniel Weiss for drawing my attention to this asymmetry.

6.   Combinations of *teplyj* with *sliškom* 'too' refer to something that is too warm for a particular purpose without being hot (cf. Section 6).

In addition to the more central temperature adjectives, Russian has a number of more restricted, domain-peripheral terms used for extremely hot and cold temperatures (cf. Table 3 and Table 8 in Sections 4 and 7). As a rule, each of the adjectives for the extremely cold and extremely hot temperatures is restricted to particular subsets of entities. For instance, *žgučij, paljaščij* 'burning, scorching, parching' are mainly used about the sun; *kipjaščij* 'boiling hot' is mainly for liquids; *raskalennyj* 'red-glowing' is mainly for surfaces; and *ledjanoj* 'ice-cold' is mainly about water, air, wind, fingers, hands and feet. There are also regular syntactic means for the expression of intensification and attenuation, as well as intensifying (*-jušč, -ejš*) and attenuative (*-ovat*) suffixes. Their uses may provide further insights into the meanings and status of the different temperature adjectives. For instance, *xolodnyj, teplyj* and *žarkij* co-occur with the attenuative suffix *–ovat*, while this is excluded for *gorjačij* and *proxladnyj*.

Proxladnyj is the most problematic temperature adjective when it comes to its status as domain-central vs. domain-peripheral. On the one hand, it satisfies the main criterion for being domain-central in view of its broad applicability. On the other hand, it is much less frequent than the other four domain-central terms, as can be seen in Table 2, which summarizes the frequencies of the central temperature adjectives in the Russian National Corpus[7] in all their uses, both concrete and extended. The table shows that the two 'hot' adjectives together have the highest frequency, followed by *xolodnyj* 'cold' and then by *teplyj* 'warm'. *Proxladnyj* 'cool' has a significantly lower frequency, constituting 11% of the frequency of *xolodnyj* 'cold' and 35.5% of the least frequent of the two 'hot' adjectives, *žarkij* 'hot non-tactile'. Moreover, it seems to have a less independent place in the Russian temperature system than the other four domain-central terms: its meaning is more or less similar to that of *xolodn-ovatyj*, i.e. *xolodnyj* 'cold' with the attenuative suffix, and it hardly participates in any regular antonymic relations with any of the other temperature terms. This shows that the distinction between domain-central and domain-peripheral temperature terms is relatively fluid. Furthermore, as will be discussed in Section 8, the 'cool' vs. 'cold' distinction is even more problematic in most of the other Slavic languages.

---

7.   http://ruscorpora.ru/index.html accessed on January 6th 2017

Table 2. The frequencies of the central Russian temperature adjectives in the Russian National Corpus (http://ruscorpora.ru/index.html, accessed on January 2nd 2017)

Temperature term	Number of occurrences	Relative frequency	
gorjačij 'hot'	35,777	24.44%	34.33%
žarkij 'hot'	14,484	9.89%	
teplyj 'warm'	44,331	30.29%	
xolodnyj 'cold'	46,636	31.86%	
proxladnyj 'cool'	5,148	3.51%	
Totals	**146,376**	**100%**	

The next sections will take the Russian temperature-term system as the point of departure for comparing temperature-term systems across the Slavic languages. The discussion will concern both the actual words/stems and the organization of the systems, raising issues related to semantic shifts and lexical replacement.

4. Slavic temperature adjectives for the warming zone ('warm', 'hot', 'extremely hot'): forms and origin

Table 3 lists 'warm', 'hot' and 'extreme hot' adjectives across Slavic, for the moment glossing over the meaning differences among them; these will be considered in the next two sections.

Table 3. Temperature adjectives for 'warm', 'hot', and 'extremely hot' across Slavic

	Language	'warm'	'hot'	'extremely hot'
East-Slavic	Byelorussian	*cëply*	*garačy* [*žarki* (coll.)]	*spjakotlivy* 'sultry' *spjakotny* 'sultry' *paljučy* 'burning' *pryparny* 'steamy, sultry'
	Russian	*teplyj*	*gorjačij* *žarkij*	*obžigajuščij* 'burning' *raskalennyj* 'red-glowing' *znojnyj* 'very hot & sunny' *paljaščij* 'burning' (about the sun) *pekuščij* 'baking' (about the sun)
	Ukrainian	*teplyj*	*harjačyj* [*žarkyj*]	*spekotlyvyj* 'sultry' *paljučyj* 'burning' *parkyj* 'steamy, sultry' *parnyj* 'steamy, sultry' *varkyj* 'cooking' *spečnyj* 'stinging, baking' *pekučyj* 'stinging, baking'

Table 3. (*continued*)

	Language	'warm'	'hot'	'extremely hot'
West-Slavic	Czech	*teplý*	*horký*	*vařící* 'boiling, boiling hot' *žhavý* 'glowing, red-hot'
	Polish	*ciepły*	*gorący*	*upalny* 'sweltering' *skwarny* 'scorching' *palący* 'burning' *piekący* 'stinging, baking' *parny* 'steamy, sultry' *grzany* 'heated / mulled'
	Slovak	*teplý*	*horúci*	*vriaci* 'boiling, boiling hot' *žeravý* 'glowing, fiery, molten'
	Lower Sorbian	*šopły*	*górucy*	*palecy* 'burning' *parjecy* 'boiling, boiling hot' *smalaty* 'burning, torrid, singeing' *warjecy* 'boiling, boiling hot' *linjecy* 'boiling, boiling hot' *žaglecy* 'glowing, red-hot' *žagły* 'glowing, red-hot' (old) *tešny* 'sultry'
	Upper Sorbian	*ćopły*	*horcy*	*smalaty* 'burning, torrid, singeing' *sapacy* 'burning, aflame, blazing' *warjacy* 'boiling, boiling hot' *parjaty* 'boiling, boiling hot' and 'steamy, sultry' *žehliwy* 'glowing, red-hot' *žahły* 'glowing, red-hot' *tužny* 'sultry'
South-Slavic	Bulgarian	*topăl*	*gorešt*	*nažežen* 'burning, glowing' *žarăk* 'glowing, scorching' *parešt* 'burning' *žežăk* 'burning'
	Croatian	*topao*	*vruć*	*vreo* 'boiling hot' *žarki* 'torrid, scorching'
	Macedonian	*topol*	*žežok* [*vrel*]	*vrel* 'boiling' *vžešten* 'heated' *gorešt* 'burning' *vruč* 'boiling hot' (dial.)
	Serbian	*topao*	*vruć*	*vreo* 'boiling hot' *žarki* 'torrid, scorching'
	Slovene	*topel*	*gorek*, *vroč*	*vrel* 'boiling hot' *žgoč* 'burning' *pekoč* 'burning, baking' *goreč* 'burning' *žareč* 'glowing, scorching' *razžarjen* 'red-glowing' *pripekajoč* 'baking' (about the sun)

Legend: [] = a marginal alternative for a particular temperature value

As is evident from Table 3, Slavic languages align in expressing 'warm' by one and the same stem. For 'hot', there is a certain variation: most of the Slavic languages have cognates of the Russian *gorjačij*, but the South-Slavic languages Serbian, Croatian and Slovene have *vruć/vroč*, and Macedonian has *žežok*. For the extremely hot temperatures the variation is still greater due to the fact that each term normally has quite a limited distribution, being restricted to particular subsets of entities (cf. the end of Section 3 for the Russian examples). The recurrent etymological sources for the terms for 'warm', 'hot' and 'extremely hot' are various actions associated with fire including actions involved in food preparation (burning, parching, scorching, glowing, baking, roasting), and states resulting from excessive heating (red-glowing, baked, roasted, etc.). The difference between actions and states corresponds most frequently to the difference between the erstwhile active and passive participles. The other common group of etymological sources involves clearly observable manifestations of heating water, such as boiling and cooking, and vapour as the result of heating (the latter gives rise to both deverbal and denominal adjectives).

The data on the etymological sources for the Slavic warming adjectives are summarized in Table 4.

Table 4. Etymological sources for the Slavic temperature adjectives for the warming zone

Value	Stem	Origin and cognates	Distribution
'warm'	*tep-ьl (PSl)	< *tep+-ьl; PIE root *tep- / *top) 'to be warm', also found in Old Indic *tápati (-te)* 'to warm, to heat, to burn', *tapas* 'heat, torment', Latin *tepeō, tepēre* 'to be tepid, warm', *tepidus* 'warm, mild', Hittite *tapašša-* 'heat' and Irish *ten* (< *tepn) 'fire'. The same root with another vowel in the PSl form *topiti* 'to fire, to heat' (Černyx 1994, 2: 236–237)	Everywhere in Slavic
'hot'	*gore (PSl) *gēr (PSl)	< Present participle from *gorěti 'burn, flame, shine'. The same root with another vowel (evidenced by the palatalization effects in such forms as *žarkij*). Descendants of PIE *gʷher-: *gʷhor etc. 'hot, warm' and related to Greek *thermos* 'hot, warm', Latin *formus* 'warm', Armenian *džerm* 'warm' (nowadays mainly in metaphors), Old Indic *gharmáḥ* 'heat, hot' etc. (Černyx 1994, 1: 524, 544)	East-Slavic (Byelorussian Russian, Ukrainian), West-Slavic (Polish, Czech, Slovak, U. and L. Sorbian), Bulgarian (South-Slavic)

Table 4. (*continued*)

Value	Stem	Origin and cognates	Distribution
	vьrěti (PSl)	'to boil (intr.), to seethe', e.g. Czech *vařít (se)* 'to cook, to boil) or Old Church Slavonic *variti, varъ* 'heat'. The pan-Slavic forms find close correspondences in Lithuanian *vìrti* and Latvian *virt* 'to cook, to boil'. Unclear to what extent these words are the result of an independent development from PIE or have been borrowed from Slavic (Černyx 1994, 1: 135).	Serbian, Croatian, Slovene (South-Slavic)
	žegti (PSl)	'to burn', e.g. *žeč* in Russian, *žgati* in Slovene; PIE *$dheg\ ^{u}h$-/*$dhog\ ^{u}h$- 'to burn, scorch, bake'. Multiple cognates across IE, including Lithuanian *dègti* 'to burn', Avesta *dažaiti* 'burns', and even Old Prussian *dagis* 'summer' and English *day* (Černyx 1994, 1: 301).	Macedonian (South-Slavic)
'extr. hot'	many	deverbal adjectives (diachronically participles) and denominal adjectives, recruited from different sources.	

5. Temperature-frame distinctions across Slavic

As shown in Section 3, the Russian temperature-term system is sensitive to distinctions among the frames of temperature evaluation, both in its morphosyntactic patterns and in its lexical choices, the latter confined to the 'hot' zone. Several issues of relevance for cross-Slavic comparison are worth mentioning in this context.

To start with, the different morphosyntactic patterns summarized in Figure 3 for Russian are not restricted to temperature. The impersonal constructions are regularly used in other predications referring to ambient and experiencer states, and there are many adjectives beyond temperature ones that participate in the same contrasts among property vs. ambient vs. experiencer predications, e.g. *temnyj* 'dark', *strašnyj* 'scary', *krasivyj* 'beautiful', *ploxoj* 'bad', etc.; see (5).[8]

---

8. As noted by Zimmerling (2018: 47), some predicatives can be used in impersonal ambient constructions, but not in experiencer ones, e.g., *solnečno* 'sunny', *pyl'no* 'dusty', while the reverse case is not attested. This is also true for some of the domain-peripheral temperature terms, both in Russian and in other Slavic languages.

(5)   Russian (East-Slavic): 'bad' as predicate

    a.   Property predication
       P'es-a   *byl-a*   *plox-aja/oj.*
       play-NOM.SG was-F.SG bad-NOM.F.SG/INS.F.SG
       'The play was bad.'

    b.   Non-referential ambient predication
       Tam / Včera   *byl-o*   *plox-o.*
       there / yesterday was-N.SG bad-PRED
       'It was bad there / yesterday.'

    c.   Personal-feeling predication
       Mne /Pet-e   *byl-o*   *plox-o.*
       I:DAT /peter-DAT was-N.SG bad-PRED
       'I / Peter felt bad.'

The Russian temperature-term system therefore "benefits" from the more general morphosyntactic distinctions of the language. In other words, the three main constructions summarized in Figure 3 (Cx_1, Cx_2 and Cx_3) can be seen as instances of the more general constructions of Property predication, Non-referential ambient predication and Personal-feeling predication, and the temperature constructions inherit their properties. The lexical distinction between the two 'hot' terms is, on the other hand, inherent to the temperature system itself. Importantly, the two systems, i.e., the morphosyntactic and the lexical one, do not always work in tandem in the sense that the distinctions set up by the one may differ from those set up by the other one, as will be shown immediately below.

First of all, as is evident from the examples in (4), the non-tactile hot *žarkij* covers **all** ambient evaluation (both non-referential and quasi-referential), as opposed to the impersonal constructions which are restricted to non-referential ambient contexts. The relevant question here is which entities count as "ambient" and collocate with *žarkij*. These include weather and climate, periods of time, geographic spaces and indoor spaces, sources of heat, clothes, etc. Interestingly, a few entities, such as heat conductors (air, wind), sources of heat (sun, fire, heater) and clothes, are compatible with different construal of their temperature properties – both tactile and ambient. The combinations *gorjačij vozdux* and *žarkij vozdux* ('hot air') pertain to the same extra-linguistic situation, but describe slightly different experiences and give rise to different images, evoking the image of the "burning" effect of the air on one's skin vs. its effect on a person as a whole (e.g., how difficult it is to breathe or how wonderful it is to do without heavy clothes) (for details, see Koptjevskaja-Tamm & Rakhilina 2006). This double behaviour of words for 'air' and 'wind' makes sense. Air is of course ambient, but it also comes in contact with our skin, even though this is normally not registered. However, this contact can sometimes become more salient, for instance, when the skin itself is affected by

very low or by very high temperatures. The same may go *mutatis mutandi* for fire, sun and other sources of heat (they may keep you warm or make you hot, but may also burn you).[9]

Persons also have an exclusive status here, given that they are the primary target of personal-feeling (and of bodily-symptom) temperatures, but are also the primary experiencers of any other temperatures; in addition, their own bodies may be evaluated in a tactile way. The examples in (4d) and (6) illustrate how lexical and morphosyntactic choices interact to discriminate among these different meanings, in this case, depending on the different roles of one and the same person, Petja, in temperature-related situations. Examples (4d) and (6a) are straightforward since they combine each of the two 'hot' lexemes with their regular predicative construction: the non-tactile *žarkij* with the impersonal experiencer construction for personal-feeling temperature in (4d), and the tactile *gorjačij* with the canonical property predication for tactile temperature in (6a). In (6b), however, the tactile-temperature lexeme is used in the experiencer construction, with the result that what is evaluated is no longer the experiencer's "internal state", but something external, the temperature of which can be estimated in a tactile way. This sentence would be appropriate if the water with which Petja is washing his hands is too hot, or when he has touched a very hot surface. However, the combination of the non-tactile 'hot' with the standard property-predicational construction does not make sense for human beings (6c).

(6) Russian
    a.   Tactile (stimulus)
        *Petj-a   sovsem   **gorjač-ij**.*
        Petja-NOM completely hot_tact-NOM.M.SG
        'Petja is hot to touch = Petja has a fever.'
    b.   Tactile (experiencer)
        *Pet-e   (sliškom) **gorjač-o**.*
        Petja:DAT (too)   hot_tact-PRED
        'It (something) is too hot for Petja to touch'.
    c.   **Petj-a   (sovsem)   **žark-ij**.*
        Petja-NOM completely hot_nontact-NOM.M.SG

The connection between Examples (6a) and (6c) and the constructions summarized in Figure 3 is uncomplicated. Example (6a) is compatible with the Tactile temperature predication $Cx_{1.1}$ in that it involves a concrete non-ambient entity (Petja) and the temperature adjective *gorjačij* can refer to tactile temperature. Example (6c), on

---

9. While some languages may also treat water as ambient in their temperature systems, this is not done in Russian.

the other hand, is not compatible with either $Cx_{1.1}$ or $Cx_{1.2}$. In particular, $Cx_{1.1}$ is excluded because *žarkij* cannot refer to tactile temperatures, while $Cx_{1.2}$ is excluded because Petja is not an ambient entity. The meaning of (6b) is slightly less straightforward, but has to do with the fact that constructions with predicatives come in several structurally and semantically different types, with the concomitant multifunctionality of the Dative NPs involved in them (cf. Letuchij 2017). In (6b), the predicative describes the properties of an unnamed entity and the Dative NP refers to the person evaluating them, cf. other similar examples like *mne (sliškom) sladko/ mjagko* 'this is (too) sweet/soft for me' (e.g., when tasting a dish or testing a bed).

Morphosyntactic distinctions among frames of temperature evaluations, similar to those in Russian, are found everywhere in Slavic languages. This is illustrated in (7) for Polish (West-Slavic) and in (8) for Serbian (South-Slavic).

(7) Polish (West-Slavic; Marcin Włodarczak p.c.)
 a. Tactile:
   Wod-a   jest ciepł-a.
   water-NOM.SG is   warm-NOM.F.SG
   'The water is warm.'
 b. Non-referential ambient:
   W pokoj-u   był-o   ciepł-o.
   in room-LOC.SG was-N.SG warm-PRED
   'It was warm in the room.'
 c. Personal-feeling:
   Był-o   mi   ciepł-o.
   was-N.SG I:DAT warm-PRED
   'I felt warm.'

(8) Serbian (South-Slavic; Katarina Rasulic p.c.)
 a. Tactile:
   Vod-a   je topl-a.
   water-NOM.SG is warm-NOM.F.SG
   'The water is warm.'
 b. Non-referential ambient:
   Danas je topl-o.
   today is warm-PRED
   'It is warm today.'
 c. Personal-feeling
   Meni je topl-o.
   I:DAT is warm-PRED
   'I feel warm.'

The impersonal ambient and personal-feeling constructions involve a predicative which is typically systematically related to the corresponding adjective (cf. the explanation preceding (2) and fn. 4). Sometimes the relation is less straightforward and accompanied by peculiarities. For instance, the predicative form for 'cold' in Czech, Slovak and Upper and Lower Sorbian has the same form as the noun *zima* 'winter', related to the 'cold' adjective in Sorbian (*zymny*), but not in Czech or Slovak (*studený*). At least Czech shows an interesting alternation between ambient and personal-feeling temperature predication involving the more or less canonical clause structure with the nominal expression as the subject (9a) and the impersonal construction with the predicative (9b). While the nominal expression in (9a) allows adjectival modification and requires gender agreement on the copula (being therefore a version of the quasi-referential construction, here also used for personal-feeling temperatures), the predicative expression in (9b) (marginally) allows adverbial modification and has a default (neuter) gender form of the copula.

(9) Czech (West Slavic; Viktor Elšik p.c.): personal-feeling predications for 'cold'
   a. *Zima* as noun:
       *Byl**a**   mi   (**velk-á**)   zim-a.*
       was.F.3SG I.DAT big-NOM.F.SG cold-NOM.SG
   b. *Zima* as: predicative
       *Byl**o**   mi   (**moc**) zima.*
       was.N.3SG I.DAT very   cold.PRED
       'I felt (very) cold.'

However, of all the Slavic languages, it is only modern Russian that has an obligatory lexical distinction between the tactile and non-tactile 'hot' adjectives. As shown in Table 4 in Section 4, this distinction involves two adjectives that go back to the same Indo-European stem *$*g^{u}her$-*, but with different vowels. *Gorjačij* is originally a present participle from the proto-Slavic verb *$*goręti$* 'burn, flame, shine', with cognates across Slavic. *Žarkij*-cognates are found in some other Slavic languages for extremely hot temperatures. Their restricted uses stem from their close associations with *žar* 'embers' ('red-hot', 'glowing', etc.), with 'sun' and 'fire' as the leading collocates.

For instance, in Serbian (South-Slavic), according to Rasulic (2015: 264), *žarki* 'extremely hot, torrid' is primarily used for high ambient temperatures, typically though not necessarily unpleasant, but only in attribution and in quasi-referential predication with expressions referring to climate, weather, summer, day, desert, sun, etc. It can occasionally apply to very few extremely hot entities within the tactile frame (e.g., lava and sand). And, most importantly, it is forbidden in the impersonal constructions dedicated to personal-feeling and non-referential ambient

predication illustrated in (8b–c), cf. *Žarko mi je 'I feel hot' and *Danas je žarko 'It is hot today'.

But even the closest relatives of Russian, Byelorussian and Ukrainian, lack the systematic frame-related contrast for 'hot'. In both languages *žarki/ žarkyj* is restricted to the non-tactile 'hot', but is only an optional alternative to the much more frequent *garačy/ harjačyj*.

In Byelorussian *garačy* is frame-neutral and occurs in all the morphosyntactic patterns characteristic for the different frames, cf. the tactile uses in *garačaja vada* 'hot water', *garačyja ruki* 'hot hands' vs. the ambient uses in *garačy dzen'* 'hot day' or *garačyja krainy* 'hot countries'. *Žarki*, considered colloquial (Tatiana Antontchik p.c.), is only used in ambient contexts, e.g. *žarkae polymja* 'hot flame', *žarki dzen'* 'hot day'. However, according to the historical dictionary of Byelorussian (HSBM), it has been around for a few centuries, with examples dating to the 16th century.

Likewise in Ukrainian, according to Kryvenko (2015: 313–317, cf. also SUM),[10] *harjačyj* is the generic adjective for 'hot', used for all the temperature frames, while *žarkyj* is restricted to ambient and slightly higher temperatures (cf. Examples (10) and (13b) in Section 6).

(10)   Ukrainian (East-Slavic; Anna Kryvenko p.c.)
    a.   Tactile:
        harjača/ *žarka voda
        'hot water'
    b.   Non-referential ambient:
        Nadvori harjač-e/ žark-o.
        outdoors "harjač"-PRED/"žark"-PRED
        'It is hot /very hot outdoors'.
    c.   Personal-feeling:
        Meni harjač-e/ žark-o.
        I:DAT "harjač"-PRED/"žark"-PRED
        'I feel hot/very hot.'

There are a few contexts where *žarkyj* cannot be replaced by *harjačyj* (SUM, Kryvenko 2015: 313–315), but on the whole it has a significantly lower frequency than *harjačyj*.

Table 5 sums up the data on the properties of *žarkij* and its cognates in Serbian, Ukrainian and Russian.

---

10. http://sum.in.ua

Table 5. "Žarkij" vs. "gorjačij" in Russian and their correspondences
in Serbian and Ukrainian

Language	Term	Value	Frames of temperature evaluation			
			'tactile'	'quasi-ref. ambient'	'non-ref. ambient	'personal-feeling'
Serbian	*žarki*	'extremely hot'	–*	+	–	–
	vreo	'boiling hot'	+	+	+	±
	vruć	'hot'	+	+	+	+
Ukrainian	*žarkyj*	'torrid, hot'	–	+	+	+
	harjačyj	'hot'	+	+	+	+
Russian	*žarkij*	'hot'	–	+	+	+
	gorjačij	'hot'	+	–	–	–

Note.

* *Žarki* can occasionally apply to a few extremely hot entities within the tactile frame, e.g., lava and sand, but not to geysers or rocks

The data in Table 5 can be interpreted as possible steps in the establishment of the lexical opposition between the tactile (*gorjačij*) and the non-tactile (*žarkij*) 'hot' in Russian, as follows:

a. The non-tactile term for 'hot' is recruited among the domain-peripheral terms for 'extremely hot', restricted to ambient entities (as in Serbian).
b. It gradually becomes less "extreme" in its temperature value and expands its application to further non-tactile frames (quasi-referential ambient and personal-feeling temperatures), where it still competes with the general term for 'hot', as in Ukrainian.
c. Finally, the non-tactile 'hot' ousts the erstwhile general term for 'hot' from the non-tactile contexts, leading to the complementary distribution between the two terms with the same temperature value, as in Russian.

But even in Russian itself, the radical distinction between the tactile and the non-tactile hot adjectives seems to be relatively recent. The Dictionary of the 18th century's Russian (SRJa XVIII) gives *gorjačij* and *žarkij* as synonyms in reference to weather, climate, and locations, as exemplified in (11).

(11) a. *V amerikanskix   ostrovax   klimat …   očen'*
         in American:PL.LOC island:PL.LOC climate:NOM.SG very
         gorjač'.
         "gorjač'":NOM.M.SG.
         'In the American islands the climate… is very hot.'.

(Sl. Komm. II 470 in SRJa XVIII)

b. *Spat', otvoriv okny, ne menĕe opasno.*
Sleep:INF open:CVB.PRF window:PL.ACC NEG less dangerous:PRED
Ètogo nikogda ne nadobno dĕlat', daže v
this:GEN never NEG need do:INF even in
gorjač-ajš-uju *lĕt-nj-uju por-u.*
"gorjač"-SUPERL-ACC.F.SG summer-A-ACC.F.SG time-ACC.SG
'It is no less dangerous to sleep with open windows. One should never do this, not even during the **hottest** summer time.'

(Dom. Leč. I 389 in SRJa XVIII)

c. *Sie vlijanie ves'ma čuvstvitel'no v ves'ma*
this:NOM.N.SG influence:NOM.SG very sensitive:NOM.N.SG in very
žark-ix *i ves'ma xolodn-yx klimat-ax.*
"žark"-LOC.PL and very cold-LOC.PL climate-LOC.PL
'This influence is very noticeable in very hot and very cold climates.'

(Spb. Ž. I 103 in SRJa XVIII)

d. *I son pod tĕn'-ju drev gust-oju*
and sleep:NOM.SG under shadow-INS.SG tree:GEN.PL dense-INS.F.SG
*Prijaten v **žark-ie** čas-y*
nice:NOM.M.SG in "žark"-ACC.PL hour-ACC.PL
'And a slumber under the dense shadow of trees is nice during hot hours.'

(Lom. SS I 62 in SRJa XVIII)

6.   The 'warm' vs. 'hot' distinction in Slavic

As stated in Section 3, Russian *teplyj* 'warm' covers a narrow temperature zone, while the denotation of both *gorjačij* and *žarkij* spans from pleasant warming to unpleasant, dangerous burning temperatures. It was also mentioned that the main temperature contrast in Russian is between 'cold' and 'hot', as seen in various contexts, including conventionalized pairs of expressions involving dishes, drinks, water taps, etc. However, far from all of the Slavic languages follow this pattern. As Table 6 shows, for the two kinds of appetizers (hors d'oevres) 'cold' contrasts with 'hot' in the East-Slavic languages, and with 'warm' in the South-Slavic languages and in most of the West-Slavic languages. Polish wavers here: the more fixed and official phrase, e.g. appearing on a restaurant menu and in recipes, follows the East-Slavic pattern in contrasting 'cold' and 'hot' for food and drinks, but in other registers 'warm' seems to be preferred. 'Warm' is also contrasted with 'cold' in other conventionalized expressions, such as denoting water taps (Marta Andersson and Marcin Włodarczak p.c.).

Table 6. The main opposition in conventionalized pairs

		'cold'	'warm'	'hot'	'appetizers'
East-Slavic	Byelorussian	*xalodnyja*	*(*teplyja)*	*garačyja*	*zakuski*
	Russian	*xolodnye*	*(*teplye)*	*gorjačie*	*zakuski*
	Ukrainian	*xolodni*	*(*tepli)*	*garjači*	*zakusky*
West-Slavic	Polish	*zimne*	*ciepłe*	*gorące*	*zakąski*
	Czech	*studené*	*teplé*	*(*horké)*	*předkrmy*
	Slovak	*studené*	*teplé*	*(*horúce)*	*predjedlá*
	L. Sorbian	*zymne*	*šopłe*	*(*góruce)*	*pśedjěze*
	U. Sorbian	*zymne*	*ćopłe*	*(*horce)*	*zakuski/předjědźe*
South-Slavic	Bulgarian	*studeni*	*tepli*	*(*gorešti)*	*ordjovri*
	Macedonian	*ladni*	*topli*	*(*žežki)*	*predjaden'a*
	Serbian	*hladna*	*topla*	*(*vruća)*	*predjela*
	Slovene	*hladne*	*tople*	*(*vroče)*	*predjedi*

This cross-linguistic difference in what constitutes the main temperature opposition is confirmed by further evidence (cf. the discussion of in-between temperatures in Section 9). Firstly, my consultants for the South-Slavic and most of the West-Slavic languages agree that the main temperature opposition in their languages is between 'cold' and 'warm', whereas consultants for the East-Slavic languages agree that it is 'cold' vs. 'hot' that constitutes the major contrast. Preliminary frequency counts in the available corpora also point to this conclusion. For instance, in the Russian National Corpus (cf. Table 2 in Section 3) and in the test version of Ukrainian Text Corpus (cf. Table 10 in Section 8) the adjectives for 'hot' (*gorjačij + žarkij* in Russian vs. *harjačij* in Ukrainian) occur slightly more frequently than those for 'warm' (*teplyj*). In contrast to this, *topao* 'warm' occurs much more frequently than the two 'hot' adjectives *vruć* and *vreo* in the Non-annotated Corpus of Contemporary Serbian Language (cf. Table 11 in Section 8).[11]

The range of temperatures covered by terms for 'warm' in Czech, Slovak, Upper and Lower Sorbian, Croatian, Serbian and Slovene is fairly broad. The relevant terms can also attach various intensifiers, such as the intensifying prefix *pre-* (South Slavic) /*pře-* (Upper Sorbian) / pse- (Lower Sobian), or the intensifying particles and adverbs such as *moc* 'very' and *příliš* 'too much' in Czech, *jara* 'very' in Upper

---

[11] Figures for the main temperature hits in the corresponding corpus are: for *gorjačij + žarkij* vs. *teplyj* in Russian – 34.33% vs. 30.29%; for *harjači* vs. *teplyj* in Ukrainian – 24% vs. 22.8%; for *topao* vs. *vruć* vs. *vreo* in Serbian – 34.61% vs. 8% vs. 7.65%. The lists of the temperature terms included in the counts differ across Russian, Ukrainian and Serbian, which influences the relative shares of the targeted terms among all the terms counted.

Sorbian, *wjelgin* 'very', *pśeliš* 'too' (and even *cu* 'too' from the German *zu*) in Lower Sorbian, etc. Intensification is an additional property that distinguishes *topao/topel* in these languages from their cognate in Russian. As mentioned in Section 3, the temperature reference of *teplyj* in Russian is fairly fixed, and even the addition of intensifiers does not raise its level beyond pleasantness. An apparent exception here are combinations with the intensifier *sliškom* 'too', which most often apply to clothes and blankets that make one sweat and feel uncomfortable, but can occasionally be used for other situations when something has a higher temperature than what would be appropriate for a particular purpose (e.g., water for watering flowers). Similar problems with intensification is also the case for *ciepły* in Polish, which again shows its in-between position between the East-Slavic and West-Slavic systems. My consultants' judgements vary slightly as to how much the temperature designation of 'warm' terms can be raised by intensifiers and to what 'too warm' can refer to, apart from denoting something that is hotter than would be to one's liking, hot beyond the level of comfort, especially for a particular activity.[12] However, there is still a categorical distinction between terms for 'warm' and 'hot' in that the former are not able to refer to very high and dangerous temperatures, regardless of intensifiers.

In contrast, the South-Slavic language Macedonian seems to have taken the step of extending the reference of *topol* even further, making the 'hot' terms fairly marginal, and primarily reserving them for emphasizing the dangerous consequences of extremely hot temperatures. This is nicely demonstrated in (12), where *topol* and *žešok* refer to the same high temperature, but the latter clearly delimits it and accentuates its dangerous properties.

(12) Macedonian (South-Slavic; Liljana Mitkovska p.c.)
     Eve ti   topla   voda.   Vnimavaj,   žeška   e.
     here you:DAT warm:F.SG water.SG be.careful:IMP.SG hot:F.SG is
     'Here's some hot water. Be careful, it's very hot/boiling.'

The peculiar status of Macedonian *topol* in the Slavic context is further evidenced by the translations of the following two examples from Michail Bulgakov's Russian novel "Master and Margarita" into various Slavic languages. The data come from "ParaSol: A Parallel Corpus of Slavic and other languages", Regensburg University & University of Bern,[13] cf. Waldenfels (2006).

---

12. The variation in the consultants' judgements likely reflects cross-linguistic differences among the systems and calls for further research. For instance, my Polish consultants insist that 'too warm' always means 'too warm for a particular purpose' without implying any high temperatures per se (like in Russian), whereas this reading is less salient for the other consultants.

13. http://www.uni-regensburg.de/Fakultaeten/phil_Fak_IV/Slavistik/RPC/

(13) 'He feels hot' in Bulgakov, M. "Master i Margarita"

    a.   Russian (original):

        [*On istomlen   golod-om   i   žažd-oj.*]   *Emu*

        he   exhausted:NOM.M.SG hunger-INS.SG and thirst-INS.SG he:DAT

        žark-o.

        hot_nontact-PRED

        'He is exhausted by hunger and thirst. He feels hot.'

    b.   Ukrainian:[14]

        Iomu ***žark-o****.* / *Iomu* ***garjač-e.***

        he:DAT hot-PRED he:DAT hot-PRED

    c.   Polish:[15]

        *…****gorąc-o*** *mu.* / *Jest*   *mu* ***gorąc-o.***

        hot-PRED he:DAT be:PRS.3SG he:DAT hot-PRED

    d.   Czech:

        …a navíc   *je*   ***vedr-o.***

        and moreover be:PRS.3SG hot-PRED

    e.   Slovak:

        …a k tomu tá   ***horúčava***

        and to this   this:F.SG heat

    f.   Croatian:

        Vruć-e *mu*   *je.*

        hot-PRED he:DAT be:PRS.3SG

    g.   Serbian:

        … ***vrućina*** *mu*   *je!*

        heat   he:DAT be:PRS.3SG

    h.   Bulgarian:

        Gorešt-o *mu*   *e.*

        hot-PRED he:DAT be:PRS.3SG

    i.   Macedonian:

        …nemu mu   *e*   ***topl-o!***

        he:DAT   he:DAT be:PRS.3SG warm/hot-PRED

---

14. Two different translations.

15. Two different translations.

(14) 'A hot day' in Bulgakov, M. "Master i Margarita"

    a.   Russian (original):

Okolo četyrex čas-ov **žark-ogo** *dnja*
about four:GEN hour-GEN.PL hot_nontact-GEN.M.SG day:GEN.SG

[bol'š-aja kompanija mužčin, odet-yx v
big-NOM.F.SG company:NOM.SG man:GEN.PL dressed-GEN.PL in

štatsk-oe, vysadilas' iz trex mašin, neskol'ko ne
civil-ACC.N.SG get.out:PST.F out three:GEN car:GEN.PL somewhat NEG

doezžaja do dom-a nomer 302-bis po
arrive:PRS.CVB to house-GEN.SG number 302-bis at

Sadovoj ulic-e].
Sadovaja:DAT.F.SG street-DAT.SG

'At around four o'clock on that hot day, a big company of men in civilian clothes got out of three cars a short distance from no. 302–bis on Sadovaya Street.'

    b.   Ukrainian:

Bliz'ko četvertoi godyn-y **spekotn-ogo** *dnja*
around fourth:GEN.F.SG hour-GEN.SG sultry-GEN.M.SG day:GEN.SG

    c.   Polish:

Tegoż więc **upaln-ego** *dnia, koło*
the.same:GEN.M.SG so scorching-GEN.M.SG day:GEN.SG about

czwart-ej po południu
fourth-GEN.F.SG after noon:DAT

    d.   Czech:

Asi ve čtyři hodiny odpoledne n-oho
about at four:ACC hour:ACC.PL afternoon that-GEN.M.SG

dusn-ého *dne*
stuffy-GEN.M.SG day:GEN.SG

    e.   Slovak:

V **horúc-om** *popoludn-í okolo štvrt-ej hodin-y*
in hot-LOC.M.SG afternoon-LOC about fourth-GEN.F.SG hour-GEN.SG

    f.   Croatian:

Oko četiri sat-a tog **vrúć-eg** *dan-a*
about four hour-GEN.SG that:GEN.M.SG hot-GEN.M.SG day-GEN.SG

    g.   Serbian:

Oko četiri sat-a tog **vrel-og** *popodneva*
about four hour-GEN.SG that:GEN.M.SG very.hot-GEN.M.SG afternoon:GEN

    h.   Slovene:

Okoli štiri-h popoldne tega **vroč-ega** *dne*
about four-GEN afternoon that:GEN.M.SG hot-GEN.M.SG day:GEN.SG

i. Bulgarian:
    *kăm   četiri čas-a   na tozi **gorešt**   den*
    towards four hour-COUNT at that hot:M.SG day.SG

j. Macedonian:
    *okolu četiri-te čas ot toj mošne **topol**   den*
    about four-DEF hour of that very   warm/hot:M.SG day.SG

In Examples (13) and (14), the Russian original uses the non-tactile hot term first for personal-feeling temperature evaluation in the predicative form *žarko* and then as an attribute to the word 'day' for ambient temperature. All the translations, apart from those in Macedonian, choose terms that clearly refer to high temperatures – either the standard 'hot' term (e.g., in 13b–d, f, h for Ukrainian, Polish, Czech, Croatian and Bulgarian), one of the terms for extremely high (e.g., 14b–c, g for Ukrainian, Polish and Serbian) or 'stuffy' (14d in Czech) (ambient) temperatures, or a noun referring to heat (e.g. 13e and 13g for Slovak and Serbian). The Macedonian translation is radically different from the other Slavic languages in using *topol* – either without additional specification of the intended intensity, or together with the intensifier *mošne* 'very'.

Table 7 summarizes the properties of the cognate 'warm' adjectives *teplyj, ciepły, topao* and *topol* in Russian, Polish, Serbian and Macedonian.

Table 7. 'warm' in Russian ("teplyj"), Polish ("ciepły"), Serbian ("topao")
and Macedonian ("topol")

Language	Term	Used in the main temperature opposition	Temperature range	Relation to 'hot'
Russian	*teplyj*	–	Very narrow: warming temperatures comparable to those of a human body and/or helping to maintain it	The border fixed and obligatory; basically impossible to raise the intensity beyond the level of comfort
Polish	*ciepły*	±	Fairly narrow or relatively broad: temperatures (slightly) warmer than those of a human body and also those helping to maintain it	The border fixed and obligatory; possible to raise the level of intensity by intensifiers
Serbian	*topao*	+	Fairly broad: pleasantly/desired warming temperatures	The border fixed and obligatory; possible to significantly raise the level of intensity by intensifiers
Macedonian	*topol*	+	Fairly unlimited: warming temperatures	Used for any high temperatures; optional terms for extremely high temperatures

In other words, Russian, Polish and Serbian are similar in distinguishing between 'warm' and 'hot' in their central temperature adjectives, but differ as to where they draw the borderlines between them. By contrast, Macedonian lacks this systematic distinction.

Significantly, in older Russian, *teplyj* could refer to much higher temperatures than now and earlier seemed to be the primary antonym of 'cold', as illustrated in Example (15).

(15) Old Russian (12th century)
Posem', rekše   slavu   sjadut'   i
then   say:PTCP.PST.PL.M.NOM glory:ACC sit.down:PRS.3PL and
služivšjumu   popovi   blagoslovľšju
serve:PTCP.PST.DAT.M.SG. priest:DAT bless:PTCP.PST.DAT.M.SG
postavit'sja   im'   ot   ostavľšagosja   sočiva
put:PRS.3SG.REFL they:DAT from leave:PTCP.PRS.GEN.N.SG.REFL sochivo:GEN
na obědě,   večera   bo radi nič'so   že   variti
on dinner:LOC evening:GEN as for   nothing namely cook:INF
*podobaet',   **ne studena**   že   ego   jadjat',   n'*
be.proper:PRS.3SG NEG cold:GEN.N.SG namely it:ACC eat:PRES.3PL but
teplo.
warm/hot:ACC.N.SG
'Then, having given glory to God, they sit down and, when the priest on duty has given a blessing, the rests of the sochivo [a dish served on special religious occasions, MKT] from the dinner are put on the table (since it is not appropriate to cook in the evening), so that they should eat it not cold, but hot.'
("Ustav Studijskij", 202 ob, quoted in SPI)[16]

Examples like (15) testify to the gradual "shrinking" of the temperature range covered by *teplyj* in Russian, which has thereby become "cooler" than earlier.

7. Slavic temperature adjectives for the cooling zone ('cool', 'cold', 'extremely cold'): forms and origin

Table 8 lists 'cool', 'cold' and 'extreme cold' adjectives across Slavic, for the moment disregarding the meaning differences among them; these will be considered in the next section. Their etymological sources are summarized in Table 9.

---

16. http://feb-web.ru/feb/slovoss/ss-abc/ accessed on March 12 2017

Table 8. Temperature adjectives for 'extremely cold' – 'cold' – 'cool' across Slavic

	Language	'extremely cold'	'cold'	'cool'*
East-Slavic	Byelorussian	*le/ljadzjany* 'ice-cold' *scjudzëny* 'very cold' *marozny* 'frosty, very cold' *styly* 'frozen, stiffened' *adubely* 'stiff(ened), numb'	*xalodny* [*zimny* (coll.)]	*xaladnavaty***
	Russian	*ledjanoj* 'ice-cold' *studenyj* (coll.) 'very cold' *moroznyj* 'frosty, very cold'	*xolodnyj*	*proxladnyj*
	Ukrainian	*kryžanyj* 'ice-cold' *studenyj* 'very cold' *moroznyj* 'frosty, very cold'	*xolodnyj* [*zymnyj* (dial.)]	*proxolodnyj*
West-Slavic	Polish	*lodowaty* 'ice-cold' *mroźny* 'frosty'	*zimny*	*chłodny*
	Czech	*ledový* 'ice-cold' *mrazivý* 'frosty'	*studený*	*chladný*
	Slovak	*ľadový* ice-cold' *mrazivý* 'frosty'	*studený*	*chladný*
	Lower Sorbian	*lodzymny* 'ice-cold' *lodowaty* 'ice-cold' *zymnučki* 'intense cold' *mrozowaty* 'frosty' *marznjecy* 'frosty'	*zymny*	*chłodny*
	Upper Sorbian	*smjerćzyma* 'very cold' *lódzymny* 'ice-cold' *lodźany* 'ice-cold' *lodojty* 'ice-cold' *mrózojty* 'frosty'	*zymny*	*chłódny*
South-Slavic	Bulgarian	*leden* 'ice-cold'	*studen*	*(pro)xladen*
	Croatian	*leden* 'ice-cold' *studen* 'very cold' *promrzao* 'freezing'	*hladan*	*prohladan*
	Macedonian	*ledeni* 'ice-cold'	*studen* *laden* (coll.)	–
	Serbian	*leden* 'ice-cold' *studen* 'very cold' (arch.) *promrzao* 'freezing'	*hladan*	*prohladan*
	Slovene	*leden* 'ice-cold'	*mrzel*	*hladen*

Legend:

[] = a marginal alternative for the particular temperature value

* = the table does not include Russian *svežij* 'fresh' and its cognates across Slavic, mainly used about air and wind. The temperature meaning is secondary to such uses as 'fresh vegetables' etc.; the temperature use is also very limited

** *xaladnavaty* = an attenuative form of *xalodny* 'cold'

Surprisingly, the central 'cold' adjectives across Slavic show significant diversity in their origin. They are expressed by forms going back to four different stems and are thus represented by four cognate classes, as compared to the one cognate class for 'warm' (cf. Table 3 and Table 4 in Section 4). In several languages, the term for 'cold' goes back to the PSl *choldъ 'cold, chilliness' with the adjectival suffix *-n*. This stem is also found more or less everywhere in Slavic for 'cool', i.e. for the moderately cooling temperatures, occasionally with the prefix *pro-*. Interestingly, Russian has two versions of this stem, reflecting the split in the manifestation of *-ol-* in different groups of Slavic: the typical East-Slavic *xolod* ('cold') and the South-Slavic *xlad*, following the pattern of the numerous Church Slavic loan words in Russian ('cool').

The two West-Slavic languages Czech and Slovak and the two South-Slavic languages Bulgarian and Macedonian have forms going back to the Proto-Slavic *studъ, possibly with the adjectival suffix *-n-* added already at that stage. The other West-Slavic languages Polish and Lower and Upper Sorbian have the form related to the word for 'winter' (the coldest season) in all the Slavic languages, *zima*. Both go back to the PSl *zima 'a cold season, winter' and *zimьnъ 'cool, cold', 'pertaining to winter'. The Slovenian 'cold' *mrzel* is originally a past participle from the verb 'to freeze', recognizable across Slavic, e.g. *mёrznut'* in Russian or *mrznouti* in Czech, stemming from the PSl *mьrznǫti.

The forms for the domain-peripheral 'extremely cold', typically restricted to particular subsets of entities/frames, are again largely recognizable. The most salient motivational pattern involves forming an adjective from the word for ice, *led* in Russian and its cognates in most languages, and *kryga* in Ukrainian. We also find *studen* and the various cognates of the Slovenian *mrzel*.

Importantly, all the stems or roots manifested by the 'cooling' temperature adjectives belong to the oldest lexical stock of the Slavic languages, in the majority of cases also traceable back to proto-Indo-European. Every now and then, one or two central temperature adjectives can, in a language or a certain group of languages, also be found as more peripheral terms, i.e. being relatively restricted and less frequent than other terms. Sometimes the peripheral terms may be less frequent and stylistically marked alternatives for the expression of the same temperature values as the central terms. For instance, *laden* in Macedonian, and *zimny/zymnyj* in Byelorussian and in Ukrainian are the colloquial or dialectal alternatives for *studen* and *xalodny/xolodnyj* 'cold', respectively. The Macedonian *laden* is considered to be a loan from Serbian (Liljana Mitkovska p.c.). In Ukrainian, *zymnyj* is most probably of native, or at least of old origin, judging by the multiple attestations in folkloristic texts and in texts by numerous authors, with additional derivational suffixes listed in Grinčenko's (1907–1909) dictionary of the Ukrainian

Table 9. Etymological sources for the Slavic temperature adjectives for the cooling zone

Value	Stem	Origin and cognates	Distribution
'cold' / 'cool'	*choldъ+n(Adj) (PSl)	The origin of the Slavic root itself is not quite clear, with several hypotheses on its relation to the broader IE context. Some suggested hypotheses include the link to Proto-Germanic *kalda and Latin *gelidus* 'frozen, cold' from < PIE *gel(ə): gol(ə), or to Lithuanian *šaltas* 'cold', but both are controversial (Černyx 1994, 2: 348).	Everywhere in Slavic, but with different semantics. The main term for 'cold' in East-Slavic and in Serbian and Croatian (South-Slavic); otherwise 'cool' (but see Section 8 for details).
	*pro+ choldъ+n(Adj)		'Cool' in East-Slavic, in Serbian and Croatian.
'cold'	*zim-ьnь (PSl)	PSl *zima 'a cold season, winter' and *zimьnь 'cool, cold', 'pertaining to winter'. Zima 'winter' in all Slavic languages. Further cognates within IE: Lithuanian *žiemà* 'winter', Greek *kheĩma* 'a cold and rainy season' and Old Indic *hēmantáḥ* 'winter', < PIE *g'hei-:*g'hu 'winter, snow' (Černyx 1994, 1: 325).	Main term for 'cold' in Polish, Upper and Lower Sorbian (West-Slavic).
'cold' / 'extremely cold'	*studъ (+n, Adj) (PSl)	Cognates across Slavic, e.g. Russian *studit'* and Polish *studzić* 'to cool', Ukrainian *stuža* 'severe cold (about weather)'. Hypothetically < *PIE stem *(s)teu̯ə 'shrink', related to PIE *stā- (the same as in *stand* etc.) (Černyx 1994, 2: 214).	Main term for 'cold' in Czech and Slovak (West-Slavic) and in Bulgarian and Macedonian (South-Slavic).
'cold' / 'extremely cold'	*mьrznǫti, *morz (PSl)	The Slovene 'cold' *mrzel* < a past participle from the verb 'to freeze', e.g. *mërznut'* in Russian or *mrznout* in Czech, < PSl *mьrznǫti. Related to 'frost, severe cold' across Slavic, e.g. Russian *moroz* or Czech *mráz*, < PSl *morz. < PIE * merg'-:*morg'–:mr̥g' 'exhaust, destroy > rotten, deteriorate'. Cf. Albanian *mërdhíj (márdhur)* 'to freeze', *mardhë* 'goose-skin', *marth* 'frost', and Armenian *mrsel* 'get cold' (Černyx 1994, 1: 524, 544).	Main term for 'cold' only in Slovene. 'Extremely cold' (primarily for ambient temperatures) everywhere.
'extremely cold'	'ice' + Adj	'Ice' is *led* almost everywhere in Slavic, but *kryga* in Ukrainian.	

language[17] and in SUM.[18] Old Russian could also use *zimnij* to denote cold (as well as winter and North) in the 10th–12th centuries, even though in modern Russian it is unequivocally interpreted as having to do with winter (*zima*) (Černyx 1994, 1: 325). The Byelorussian *zimny* is, however, mentioned in ÈSBM (1985: 333) as a loan from Polish.

Cognates of what are peripheral terms for 'cold' in some languages occasionally denote 'extremely cold' in others. For instance, *studen* in Serbian and Croatian exists as an archaic and relatively marginal word, primarily used in reference to very cold temperatures, along with the central cold adjective *hladan*. In Russian, *studenyj* is also used in reference to very cold temperatures, primarily restricted to certain entities such as periods of time, wind and water, while the main cold adjective is *xolodnyj*. The historical relations between the two adjectives are complex; in earlier sources *studenъ* 'cold' is attested from the 11th century, while *xladnyj*, a clear borrowing from Church Slavonic (a South Slavic language variety), refers to 'cool, wet' (Sreznevskij 1987). Both *xolodnyj* 'cold' and *proxladnyj* 'cool', also with the Church Slavonic version of the root, are attested from the 16th century (Černyx 1994, 2: 213, 348; SRJa XI–XVII). Another instance where cognates differ in being domain-central in some languages but refer to the more extreme temperatures in others include *mrzel* 'cold' in Slovene vs. *promrzao* 'frozen/cold because of freezing' in Serbian.

8.   The 'cold' vs. 'cool' distinction in Slavic

As shown in Section 3, Russian makes a distinction between 'cold' and 'cool', generally applicable across all of the frames of evaluation. It is also mentioned there that *proxladnyj* 'cool' is much less frequent than *xolodnyj* 'cold' (*xolodnyj* is nine times as frequent as *proxladnyj* in the Russian National Corpus, cf. Table 2), and that the status of *proxladnyj* as domain-central is open to discussion. In fact, the saliency of the 'cold' vs. 'cool' distinction varies significantly across Slavic, even though Table 8 gives the impression that the distinction is robustly attested in the family. Since a detailed comparison would need a separate study, I will here limit myself to a few illustrations.

Table 10, based on Kryvenko (2016: 312), shows the frequencies of selected temperature adjectives in Ukrainian, with the words denoting 'cold' and 'cool' in bold. The Ukrainian cognate of the Russian 'cool', *proxolodnyj*, turns out to be

---

17. http://wiki.kubg.edu.ua/Зімний

18. http://sum.in.ua/s/zymnyj

extremely rare in the Ukrainian Text Corpus (KTUM). Its frequency constitutes only 1.5% compared to the frequency of the main 'cold' term *xolodnyj*. It is also very restricted in its application, modifying ambience, water and refreshing drinks (according to SUM).[19] The third adjective that can refer to low and relatively low temperatures, *zymnyj*, occurs almost 5 times as frequently as *proxolodnyj* in KTUM, but has the same temperature reference as *xolodnyj*; the difference between *zymnyj* and *xolodnyj* has to do with register and geographic distribution (cf. Section 7).

Table 10. The frequencies of the Ukrainian temperature adjectives and adverbs in the test version of Ukrainian Text Corpus (KTUM, http://www.mova.info/corpus.aspx, accessed on January 26th 2011) (Kryvenko 2015: 312)

	Temperature stem	Number of occurrences (Adj/Adv*)	Relative frequency
warming	*spe(k/č)-* 'scorching'	16 (14/2)	1%
	žar- 'hot, torrid'	21 (11/10)	1.3%
	harjač- 'hot'	385 (361/24)	24%
	tepl- 'warm'	364 (330/34)	22.8%
cooling	**proxolod-** 'cool'	**7 (4/3)**	**0.4%**
	zymn- 'chilly. cold'	**43 (33/10)**	**2.7%**
	xolod- 'cold'	**462 (403/59)**	**28.9%**
	stu(ž/d)- 'bitter cold'	22 (20/2)	1.4%
	kry(h/ž)- 'icy'	146 (146/0)	9.1%
	moroz- 'frosty'	174 (174/0)	10.9%
	Totals	**1600 (1456/144)**	**100%**

Note:

* Adv. = adverbs, which also includes the predicative forms used in impersonal (non-referential) ambient and personal-feeling temperature evaluation.

Table 11, based on Rasulic (2015: 270), shows that in Serbian the 'cool' adjective *prohladan* is also quite rare compared to the other temperature adjectives; its frequency constitutes 4.5% compared to the frequency of the main cold term *hladan* in the Non-annotated Corpus of Contemporary Serbian Language (NKSSJ). According to Rasulic (2015), *prohladan* is also quite restricted in application, being mainly used for ambient temperatures and for water in natural surroundings (lake, sea, etc.), but it is more or less excluded for tactile temperatures of other entities and for personal-feeling temperatures. Finally, its relative marginality in the system is noticeable in its "general resistance to semantic extension" (Rasulic 2015: 286), i.e. in the absence of metaphors involving *prohladan*, apart from rare cases of humorous downtoning within the cooling range. The same holds for Croatian (Daniela Katunar p.c.).

---

19. http://sum.in.ua/s/prokholodnyj

Table 11. Frequencies of the Serbian temperature terms in NKSSJ (Neetiketirani korpus savremenog srpskog jezika, Non-annotated Corpus of Contemporary Serbian Language) (after Rasulic 2015: 268)**

	Temperature stem	Number of occurrences (ADJ/ImpersADV*)	Relative frequency
warming	*žarki* 'torrid/scorching'	81	1.64%
	vreo 'boiling hot'	377	7.65%
	vruć 'hot'	394	8.00%
	topao 'warm'	1705	34.61%
	mlak 'lukewarm/tepid'	195	3.96%
cooling	**prohladan** 'cool/chilly'	**78**	**1.58%**
	hladan 'cold'	**1731**	**35.13%**
	leden 'ice-cold'	366	7.43%
	Totals	4927	100%

Notes:

* ImpersADV = impersonal (non-referential) ambient and personal-feeling temperature constructions with predicatives

** These data were retrieved from the Non-annotated Corpus of the Contemporary Serbian Language (22 mil words) in February 2010. Currently there is a much larger (122 mil words) Corpus of Contemporary Serbian, SrpKor2013.

Macedonian is again more radical than its closest relatives. The on-line Macedonian dictionary *Digitalen rečnik na makedonskiot jazik* defines *laden* as 'having a low temperature, but not very cold',[20] thus giving the impression that *laden* vs. *studen* demonstrates a version of the 'cool' – 'cold' distinction. However, my consultant Liljana Mitkovska asserts instead that the two adjectives are synonymous and the difference between them is stylistic, rather than semantic.[21] Most dictionaries give similar explanations for both adjectives, but may vary the collocations (the only fixed expression with no possible replacement is *studena vojna* 'cold war'). *Laden* is considered to be Serbian-inspired and more colloquial, but is widely spread – e.g., its use for 'cold appetizers' in Table 6. There also seems to be a generational difference between the two adjectives, with young people tending to prefer *laden*.

Table 12 summarizes the data on the 'cold' – 'cool' adjectives in Russian, Ukrainian, Serbian and Macedonian.

---

20. http://www.makedonski.info/letter/л/лавра/лаѓа#ладен/прид, accessed on April 30th 2017

21. There are indications that the distinction between *studený* and *chladný* in Czech is also primarily a matter of style and register, rather than of semantics.

Table 12. The properties of 'cold' vs. 'cool' in Russian, Ukrainian, Serbian and Macedonian

Language	Term	Value	Application	Frequency/usage
Russian	*xolodnyj*	'cold'	no restrictions	
	proxladnyj	'cool'	no restrictions	11% of *xolodnyj*
Ukrainian	*xolodnyj*	'cold'	no restrictions	
	zimnyj	'cold'	no restrictions	9.5% of *xolodnyj*; colloquial/regional
	proxolodnyj	'cool'	mainly for ambience, water, refreshing drinks; all syntactic frames allowed	1.5% of *xolodnyj*
Serbian	*hladan*	'cold'		
	prohladan	'cool'	mainly for ambience and water in natural surroundings; more or less excluded for tactile temperatures of other entities and for personal-feeling temperatures	4.5% of *hladan*
Macedonian	*studen*	'cold'	no restrictions	
	laden	'cold'	no restrictions	Serbian-inspired and colloquial, but widely used and showing signs of replacing *studen*

9. In-between temperatures: 'Lukewarm'

Several Slavic languages have acquired a special term for the in-between temperatures, corresponding to the English *lukewarm* or the German *lau(warm)*. The following example from Patrick Süskind's novel "Perfume" shows translations of *lauwarm* from the original German text into English and several Slavic languages. Czech, Polish, Serbian, Croatian and Slovene (but not Russian and Bulgarian) use a special word, different from all the other temperature adjectives. (The data come from "ParaSol", cf. Waldenfels 2006.)

(16) 'Lukewarm' in Süskind, P. "Das Parfum"
    a. German (origin):
        *Dann ließ der Marquis die Ventilatoren anhalten und verbrachte Grenouille in einen Waschraum, [wo er in Bädern von **lauwarmem** Regenwasser mehrere Stunden eingeweicht] und schließlich mit Nussölseife aus der Andenstadt Potosi von Kopf bis Fuß gewaschen wurde.*

b.   English:

*Then the marquis had the ventilators stopped and Grenouille brought to a washroom, [where he was softened for several hours in baths of **lukewarm rainwater**] and finally waxed from head to toe with nut-oil soap from Potosi in the Andes.*

c.   Czech:

kde	*ho*	*několik hodin*	*namáčeli*	*do lázní*
where	he:ACC	several hour:GEN.PL	make_wet:PST.3PL	to bath:GEN.PL

z	***vlažné***	*dešťové*	*vody*
with	lukewarm:GEN.F.SG	rainy:GEN.F.SG	water:GEN.SG

d.   Polish:

gdzie	*go*	*przez wiele godzin*	*moczono*	*w*
where	he:ACC	during many hour:GEN.PL	make_wet:PASS.PTCP.N.SG	in

kąpieli	*z*	***letniej***	*wody*	*deszczowej*
bath:LOC.SG	with	lukewarm:GEN.F.SG	water:GEN.SG	rainy:GEN.F.SG

e.   Serbian:

gde	*je*	*više*	*časova*	*proveo*	*raskiseljavajući*
where	is	more	hour:GEN.PL.	spend:PST.PTCP.M.SG	macerating

se	*u*	***mlakoj***	*kišnici*
REFL	in	lukewarm:LOC.F.SG	rainwater:LOC.SG

f.   Croatian:

u kojoj	*su*	*ga*	*namakali*	*u kupkama*
in which	are	he:ACC	soak:PST.PTCP.PL	in bath:PL.LOC

mlake	*kišnice*	*nekoliko sati…*
lukewarm:GEN.F.SG	rainwater:GEN.SG	few hour:PL

g.   Slovene:

kjer	*so*	*ga*	*več*	*ur*	*namakali*	*v kopelih*
where	are	he:ACC	several	hour:GEN.PL	soak:PST.PTCP.PL	in bath:PL.LOC

z	***mlačno***	*deževnico…*
with	lukewarm:INS.F.SG	rainwater:INS.SG

h.   Bulgarian:

kăde-to	*njakolko*	*časa*	*go*	*kisna-xa*	*văv vani*
where-DEM	several	hour:COUNT	him	soak.IPV-3PL.AOR	in bath:PL

s	***xlad-k-a***	*dăždovna*	*voda*
with	cool-DIM-F.SG	rainy:F.SG	water

i.   Russian:

gde	*ego*	*neskoľko*	*časov*	*otmačivali*	*v vannax*
where	he:ACC	several	hour:GEN.PL	soak.IPFV:PST.PL	in bath:LOC.PL

s	***teploj***	*doždevoj*	*vodoj*
with	warm:INS.F.SG	rainy:INS.F.SG	water:INS.SG

Table 13 lists the terms for the in-between temperatures ('lukewarm') in those Slavic languages that have acquired them, together with their etymological sources. As the table shows, the 'lukewarm' temperature terms across Slavic manifest a great diversity of forms, that can be traced back to five different etymological sources.

Table 13. Adjectives for the in-between ('lukewarm') temperatures in Slavic languages

	Language	'LUKEWARM'	Origin and cognates
East-Slavic	Ukrainian	*liteplyj*	< *teplyj. li-* assumed to be of common Slavic origin and to add attenuation 'hardly, slightly' (ESUM, 3: 268).
		litnij (?)	< *lito/leto/lato* 'summer'. Cf. Russian *letnij* or Czech *letní* 'pertaining to summer'.
	Byelorussian	*letni* (coll.)	
West-Slavic	Polish	*letni*	
	Czech	*vlažný*	< *vlaha* 'liquid, wetness' (Rejzek 2001: 715). Cf. Russian (*vlažnyj*) and Bulgarian (*vlažen*) 'wet'.
	Slovak	*vlažný*	
	Lower Sorbian	*glěwki*	Probably < PSl **gliv-ъkъ/*glěv-ъkъ* related to the Ukrainian *glevkij* and dialectal Russian *glivkoj* 'similar to damp clay, not sufficiently baked, sticky', to the dialectal Russian *glev(a)* 'fish slime; mould; skin on beer etc.' and, outside of Slavic, to the Lithuanian *gléivės* 'slime, phlegm' and the Latvian *glēvs* 'soft, weak, cowardly' (Schuster-Šewc 1980: 855).
	Upper Sorbian	*liwki*	
South-Slavic	Macedonian	*mlak*	Two different hypotheses in different sources. (1) < South-Slavic *mlahav / mlohav* 'weak, soft', with the lukewarm meaning attested from the 16th century. Related to the Ukrainian *mľavyj* 'weak, sluggish, flat' and, further, to the Greek *βλάξ* 'weak', perhaps traceable to the IE root **m(e)lā* with the suffix *-ko* (Skok 1971–1974, 2: 439–440). (2) < *mláka* 'puddle' (Snoj 2016).[a]
	Serbian, Croatian	*mlak*	
	Slovene	*mlačen*	

a. http://www.fran.si/193/marko-snoj-slovenski-etimoloski-slovar/4288903/mlacen?View=1&Query=mla%c4%8den&FilteredDictionaryIds=19, accessed on December 30 2016.

The temperature range covered by such in-between terms is extremely restricted. These are the temperatures that are not sensed as either warm or cold, or are experienced as not sufficiently warm, and the same goes for the range of entities that allow such qualification. The prototypical entity here is water, as in the Süskind example, but also other liquids (such as tea or milk). When asked to describe the temperature of 'lukewarm' liquids, speakers of languages with such terms often talk about the temperature of liquids you give to babies in a bottle or the temperature of water for bathing babies; you can test the temperature by your wrist, fingers, elbows, etc.

Lukewarm water can occur both in household and in natural conditions, e.g., in a lake or in a puddle. Some languages may also apply their in-between temperature terms to non-liquid food and/or to air, wind, evenings and other ambient entities.

Slovene is a good case in point here. "The Dictionary of Standard Slovene" (*Slovar slovenskega knjižnega jezika*), compiled in the 1960's and using authentic examples, characterizes *mlačen* as 'a bit warm' (*nekoliko topel*) and provides examples like *piti mlačen čaj* 'to drink lukewarm tea', *umiti se v mlačni vodi* 'to wash in lukewarm water', *jed je samo še mlačna* 'the dish only remains lukewarm', *pihal je mlačen veter* 'a lukewarm wind was blowing' and *mlačen pomladni zrak* 'lukewarm spring air'.[22] However, not all speakers accept the last two examples, and *voda* 'water' is the main collocate of *mlačen* in the representative corpus of the Slovene language Gigafida (http://www.gigafida.net/). For 'lukewarm beer' Slovene allows both *mlačno pivo* 'lukewarm beer' and *toplo pivo* 'warm beer', even though the latter examples are twice as frequent in Gigafida, but here again languages can vary (I am extremely grateful to Monika Kavalir for the data on Slovene).

The stronghold of the 'lukewarm' terms in Slavic is found in West-Slavic and South-Slavic (except for Bulgarian). *Liteply* in Ukrainian is classified as dialectal in SUM and seems to be fairly infrequent (Galina Yavorska and Galina Zymovets p.c.). *Letni* in the sense of 'lukewarm' (in addition to 'pertaining to summer') is used in colloquial Byelorussian[23] and in the Byelorussian version of Russian (Tatiana Antontchik p.c.). *Letnij/ litnij* in reference to lukewarm liquids is attested in Russian dialects, according to the dictionary by Dal', first published in 1863–1866,[24] and in Ukrainian, as testified by the dictionaries published in the first half of the 20th century, e.g. Grinčenko (1907–1909) and RUAS (1924–1933),[25] and by occasional examples in current use. However, such uses do not seem to be very frequent (Galina Yavorska and Galina Zymovets p.c.).

The lack or the weak status of a dedicated 'lukewarm' term in East-Slavic does not come as surprise given that the narrow denotational range of its basic 'warm' term largely overlaps with that of the typical 'lukewarm' term, as evidenced by the Russian translation in (16i). Their descriptive meanings, however, differ, inasmuch as *teplyj* has a salient warming component in its semantics, as contrasted to the meaning of 'neutral, not sufficiently warm' in the 'lukewarm' term (cf. Koptjevskaja-Tamm & Rakhilina 2006 for the details). The Bulgarian case is less clear.

---

22. http://www.fran.si/iskanje?FilteredDictionaryIds=130&View=1&Query=mla%C4%8Den, accessed on December 30 2016

23. http://www.skarnik.by/tsbm/35766

24. http://slovardalja.net/word.php?wordid=14907

25. http://r2u.org.ua/s?w=літній&scope=all&dicts=all&highlight=on

10. Wrapping up the Slavic comparison

It is now time to summarize the findings of the current study. As stated in Section 1, the comparison targets both the *temperature systems as a whole* and the *forms* involved in them, with an eye to intra-genetic *constancy* vs. *variation*. These two aspects will be considered in Sections 10.1 and 10.2 respectively.

10.1 How (dis)similar are the temperature systems across Slavic?

Starting with the main temperature-value distinctions for the whole system and the domain-central temperature terms, Russian was argued to have a *four-value* system (see Table 1 in Section 3), distinguishing 'cold', 'cool', 'warm' and 'hot'. However, it may also be the only Slavic language of this type, given that the lexical distinction between 'cool' vs. 'cold' is normally not fully carried out across all of the frames of temperature evaluation (cf. the discussion and Table 12 in Section 8). As a result, several Slavic languages show a *three-value* temperature system 'cold/cool' vs. 'warm' vs. 'hot', e.g., Ukrainian, Byelorussian, Serbian and Croatian; I suspect that this is also the case in most of the other Slavic languages. Finally, Macedonian, which lacks the systematic distinction between 'hot' and 'warm', manifests a *two-value* system, distinguishing between 'cold/cool' and 'warm/hot'.

The Slavic languages with a lexical distinction between 'warm' and 'hot' differ as to where *the borderlines* between the members in each pair are drawn (cf. Table 7 in Section 6). In some (e.g., Russian, Ukrainian, Byelorussian), the term for 'warm' covers a fairly limited range of warming temperatures, comparable to that of a human body and/or helping to maintain it, while the denotation of the term for 'hot' stretches from pleasant warming to unpleasant, dangerous burning temperatures. In others (e.g., Czech, Serbian, Croatian), the 'warm' term covers a relatively broad range of pleasantly/desired warming temperatures, with the 'hot' term mainly reserved for unpleasantly/dangerous high warming temperatures. This has consequences for the lexical relations of antonymy within the temperature systems, where the main temperature opposition in conventionalized pairs of expressions, such as involving dishes and drinks is between 'cold' and 'hot' in the East-Slavic languages, but between 'cold' and 'warm' in the South-Slavic languages and in most West-Slavic languages. Polish turns out to be in-between (cf. Table 6 in Section 6).

Finally, Russian is the only Slavic language to make a systematic lexical distinction between the *frame-specific tactile* and the *cross-frame non-tactile* 'hot' categories (cf. Table 5 in Section 5). The main temperature-value systems found across Slavic are summarized in Table 14 (Polish is not included here due to insufficient data on the 'cool' vs. 'cold' distinction).

Table 14. The main temperature-value systems found in Slavic languages

	COOLING		WARMING		
	'general/ unpleasant'	'moderate'	'moderate/pleasant'		'unpleasant/ dangerous'
			comparable to human-body temperatures		
Russian	*xolodnyj*	*proxladnyj*	*teplyj*	TACTILE: *gorjačij*	
				NON-TACTILE: *žarkij*	
Ukrainian	*xolodnyj*, [*zimnyj*, dial.]		*teplyj*	*harjačyj*	
Serbian	*studen*		*topao*		*vruč*
Macedonian	*studen / laden* (coll.)		*topol*		

Moving further, we notice that all of the Slavic languages have a rich repertoire of *domain-peripheral terms* for extremely high temperatures and a somewhat more restricted one for extremely low temperatures. Domain-peripheral terms are often specialized for particular entities/frames. In addition, all of the Slavic languages have terms for *moderately cold* temperatures like 'cool; fresh', that are restricted to ambient evaluation (especially to ambient entities such as air, wind, morning), but in some languages are also applicable to water (cf. Table 12 and the discussion in Section 8). Finally, most of the Slavic languages have terms for the in-between 'lukewarm' temperatures, primarily applicable to water and some other liquids, with a possible addition of such ambient entities as air and wind (cf. Table 13 and the discussion in Section 9).

Lastly, the Slavic languages show an impressive uniformity in distinguishing among frames of temperature evaluation by morphosyntactic means. Tactile and quasi-referential ambient temperatures follow the standard construction for adjectival predicates, while non-referential ambient and personal-feeling temperatures are expressed by impersonal constructions involving the corresponding predicative (and an additional experiencer phrase in the Dative for personal-feeling temperatures). The different morphosyntactic patterns are, however, not peculiar to the temperature terms, but are more generally used for property vs. ambient vs. experiencer predications.

10.2   What is stable vs. changeable in the temperature terms across Slavic?

Shifting now to discussion of the actual *forms* and *form-meaning* pairings, we have throughout the chapter discussed several questions related to the origin, development and demise of the temperature terms across Slavic, i.e. to their *stability vs. change* (cf. primarily Sections 4 and 7). Table 15 summarizes the data on the

meanings (here limited to temperature values) and status (domain-central vs. domain-peripheral) of the most important temperature adjectives in Slavic, organized according to their etymological sources, where forms sharing the same origin are said to belong to the same "cognate class" (cf. Table 4, Table 9 and Table 13). The suggested meanings of the etymological sources are not repeated here.

Table 15. Meaning and status of the different temperature-term cognates across Slavic

Stem	Temperature value						
	'extreme cold'	'cold'	'cool'	'luke-warm'	'warm'	'hot'	'extreme hot'
tep-ьl				Uk (li-)	ALL		
gorę						ALL but Cr, Ma, Se	*Ma*
gēr						Ru[iii] Uk	*Sk, Cr, Se*
variti[iv]						Cr, Se, Sl	*Uk, Cz, Sk, So, Cr, Ma, Se, Sl*
choldъ: X[v] pro+X[v]		By, Ru, Uk, Cr, Se, Mc	*Bl, Cz, Pl, Sk, Sl, So*				
			Ru *Bl, Cr, Se*				
studъ	*By, Ru, Uk, Cr, Se*	Cz, Sk, Bl, Ma					
morz	*By, Ru, Uk, Pl, Sl, Cz, So, Se*	Sl					
zim-ьnь		Pl, So, (Cz),[vi] *By, Uk*					
led 'ice' +adj	*ALL but Uk, where 'ice' = kryga*						
letni(j) /litnij				*Pl, By /Uk*			
vlažný				*Cz, Sk*			

Table 15. (*continued*)

Stem	Temperature value						
	'extreme cold'	'cold'	'cool'	'luke-warm'	'warm'	'hot'	'extreme hot'
mlak /mlačen				Cr, Se, Ma, Sl			
glěwki /liwki				U/ L So			

Legend:

Bl = Bulgarian, By = Byelorussian, Cr = Croatian, Cz = Czech, Ma = Macedonian, Pl = Polish, Ru = Russian, Se = Serbian, Sl = Slovene, Sk = Slovak, U & L So = Upper and Lower Sorbian, Uk = Ukrainian

the names of the languages where the term is domain-peripheral are italicized

Notes:

i.   'warm' including 'hot' in Macedonian

ii. 'tactile hot' in Russian

iii. 'non-tactile hot' in Russian

iv. several different derivations from **variti*

v.   "pro+X" refers to the use of the words originating from **choldъ* with the prefix *pro*, as opposed to those that lack it (X)

vi. *zima* as the suppletive predicative variant of *studený* used in ambient and personal-feeling temperature predication (Section 5)

In what follows I will use the simple idea articulated in Dahl (2004: 263), but going back to comparative work in the 1960-ies and applied in several recent studies on lexical evolution (e.g., in Pagel et al. 2007; Holman et al. 2008) that words for more stable concepts show a greater tendency to have cognates in related languages than those for less stable concepts. Consequently, the number of cognate classes for a particular concept in related languages is an indication of the relative stability of these lexical items. Dahl's example compares the number of cognate classes for the concept 'three' with that for the concept 'girl' in 16 Romance languages. For 'three' there is only one cognate class preserved from Latin (and going much further back to Proto-Indo-European). For 'girl', on the contrary, none of the selected Romance languages have preserved the Latin word *puella*; what is even more interesting, most of the words are not related to each other, meaning that there must have been many separate replacement events. This, in turn, indicates that words for the concept 'girl' are historically much less stable than those for 'three'. Based on Table 15, the issue of lexical *stability* vs. *change* can be approached from two different angles – starting from the *meanings* or from the *forms*.

Taking *meanings* as a point of departure, the question is to what extent a given meaning is rendered by the same cognate classes across Slavic and whether a given meaning differs in the number of cognate classes involved in its expression across the family. Terms for both 'warm' and 'cool' go back to one proto-Slavic root each, but the latter involves three different formations, while the concepts of 'hot' and

'cold' are represented by three and four cognate classes respectively. It makes less sense to count the number of shared cognates for extreme temperatures, given that languages often have a richer repertoire of those (due to the semantic specialization of the terms). The in-between zone is, however, interesting. In spite of the same restricted application of the corresponding terms, it is populated by five different cognate classes.

This can be summarized in the lexical stability hierarchy seen from the meanings, as in Figure 4, where 'x' > 'y' indicates a greater relative stability of words for the concept 'x' as compared to those for the concept 'y'.

Temperature value		'warm' > 'cool'		> 'hot' >	'cold' >	'lukewarm'
Number of cognate classes	1	1–2 (one root with / without the prefix *pro*-)		3	4	5

Figure 4. Lexical stability hierarchy – starting from the meanings

Examining the data now from the perspective of the *forms*, i.e. the cognate classes themselves, we ask which cognate classes have developed as domain-central versus domain-peripheral adjectives, and whether the meanings have remained constant. The most frequent temperature-term adjectives across Slavic correspond to the Russian domain-central terms *teplyj* 'warm', *gorjačij* 'hot (tactile)' and *žarkij* 'hot (non-tactile)', *xolodnyj* 'cold' and *proxladnyj* 'cool', and to the Russian domain-peripheral terms for 'extreme cold' *studenyj* and *ledjanoj*. Being (quasi-) Pan-Slavic, the adjectives still vary in their meaning and status. While the cognates of *teplyj* and *ledjanoj* have the same status (domain-central for the former, and domain-peripheral for the latter) and more or less similar meaning across the family, the other three are subject to greater variation here.

Proto-Slavic **tepl* is used for the domain-central 'warm' terms in *all* of the Slavic languages. Its meaning/denotational range is, however, subject to variation across the family, depending on whether or how it is delimited with respect to higher temperatures (cf. the discussion on Macedonian vs. Serbian vs. Russian/Ukrainian and Table 14 above); and, partly, on whether there is an in-between 'lukewarm' term in the system.

Likewise, the domain-peripheral 'extreme cold' terms in *all* of the Slavic languages include adjectives derived from the word for 'ice', namely *led* 'ice' (Russian) and its cognates in most of the cases, and *kryga* in Ukrainian.

Things are somewhat more complicated with the three other (quasi-)Pan-Slavic cognate sets.

Forms going back to PSl *choldъ* 'cold, chilliness' with the adjectival suffix –*n*, attested in all the Slavic languages, manifest significant variation in both meaning ('cold' vs. 'cool') and status (domain-central vs. domain-peripheral). They correspond to the domain-central term for 'cold' in some languages, but to the domain-peripheral term for 'cool' in others. In a few languages, the same cognate with the prefix *pro-* is used for 'cool', with two different temperature terms in the same language going back to the same PSl stem. The examples include the domain-central terms for 'cold' vs. 'cool' in Russian, the domain-central term for 'cold' and the domain-peripheral term for 'cool' in Croatian and Serbian, and the two alternative domain-peripheral terms for 'cool' in Bulgarian.

Similarly, forms going back to Proto-Slavic *studъ* are domain-central terms for 'cold' in the West-Slavic languages Czech and Slovak and in the South-Slavic languages Bulgarian and Macedonian, whereas in most (if not all) of the others they are stylistically flavoured (archaic, dialectal) and domain-peripheral for 'extreme cold'.

Finally, the descendants of PSL *gorę* 'burning, flaming, shining' or other derivations from *gorěti* 'burn, flame, shine' function as domain-central terms for 'hot' in most of the Slavic languages, with the exception of the South-Slavic languages Croatian, Macedonian and Serbian. Several Slavic languages (including Croatian and Serbian) also have a domain-peripheral term for extremely hot ambient temperatures going back to the same root with another vowel (PSI *gēr*), which in Russian has become the domain-central term for 'non-tactile hot'.

Most of the forms listed in Tables 4 and 9 in Sections 4 and 7 have clear cognates within the broader Indo-European family, with the status of the PSl *choldъ* 'cold, chilliness' being somewhat less clear. It is also not quite clear to me to what extent the adjectives for 'extremely cold' derived from 'ice' share their origin rather than being (partly independent) innovations, particularly given the non-uniformity of the derivational suffixes used in them. But with these provisions, the stems built on *tepl*, *gorę*/*gēr*, *choldъ*, *studъ* and *led* 'ice' are indeed very stable across Slavic, where they were inherited from Proto-Indo-European. The other cognate classes in Table 15 are much more restricted in their frequency across the Slavic temperature adjective systems (even though they may occur in other parts of the languages vocabularies). The cognate classes themselves can therefore be arranged according to their stability across the Slavic temperature adjective systems, as shown in Figure 5, where *x* > *y* indicates a greater relative stability of cognates of x than those of y.

cognate classes	*tep-ьl	> *choldь; *led	> *gorę; *studъ	> *gēr; *morz; *variti	> *zim-ьnь; letni(j) /litnij; vlažný; mlak /mlačen; glěwki /liwki
occurrence across Slavic	all Slavic, comparable meaning (but not identical in range)	all Slavic, (slightly) different meanings (for *choldь), or almost all Slavic (for *led)	domain-central in some; domain-peripheral in several languages across the family	domain-central in one or a few languages belonging to the same sub-family; domain-peripheral in several languages	restricted to one or a few languages closely related (within a subfamily) and/or in close contact with each other

Figure 5. Lexical stability hierarchy – starting from the forms

It has been suggested by Frans Plank (2010), that "[b]asic[26] temperature terms are unusually pertinacious. Typically, they are passed on essentially unchanged and with essentially no vocabulary turn-over across hundreds of generations of grammar and lexicon acquirers for thousands of years". The combined perspectives (starting from meaning and from form), also with close attention to the systems as a whole, help us to evaluate and refine this claim.

First of all, it is true that most of the domain-central temperature terms are traceable to proto-Slavic and even to proto-Indo-European, but there are interesting differences in their stability, seen from both the meaning and the form angles. Also, the Slavic systems testify to several kinds of developments in the lexical systems. While some of the changes (e.g., the development of *choldь and *zimьnь) are less clear to me, most of the others can be summarized as follows:

– Developments from the periphery of the temperature domain to the centre
  – *replacement*
  In Croatian, Serbian and Slovene the adjectives *vrúć / vroč* coming from 'to boil, to cook' and earlier denoting 'extremely hot' (most probably restricted to liquids) have developed into the domain-central adjectives for 'hot', most probably replacing the descendants of the PSl *gorę with cognates for 'hot' in most of the other Slavic languages (cf. Table 4).

  Slovenian *mrzel* for 'cold' may also be a replacement of the stems going back to *studъ (+n, Adj) or *choldь+n(Adj) by a form earlier restricted to 'extremely cold' for ambient temperatures, but the details are less clear here (cf. Table 9).

---

26. Plank's "basic" terms are close to "domain-central" temperature terms as defined in this paper. The differences between the two terms are not relevant here.

– *new distinctions*

In Russian, the adjective *žarkij*, earlier having a restricted application to 'extremely hot, ambient' temperatures in certain contexts (still present in a few languages), has become the domain-central term for both personal-feeling and ambient 'hot'. As the result of this development, Russian has acquired a lexical distinction between the older 'hot' word *gorjačij*, now reserved to *tactile* 'hot', and the *non-tactile* 'hot' *žarkij* (cf. Table 5).

– Developments from the centre to the periphery: marginalisation

*studъ was retained as a domain-central term for 'cold' in Czech, Slovak, Bulgarian and Macedonian but became a domain-peripheral ('extremely cold') term in Byelorussian, Russian, Ukrainian, Croatian, Serbian.

– Restructuring of the oppositions

Slavic languages differ in their main temperature contrast, i.e. as to whether the main opposition is between the words for 'cold' and for 'warm' or between those for 'cold' and 'hot', which, in turn, indicates that the languages differ in the range of temperatures covered by the words for 'warm' vs. 'hot' and in how obligatory the distinction is (cf. Table 6, Table 7, Table 14 and Section 6).[27]

– Lexical innovations

Several Slavic languages have introduced a term for 'lukewarm' and thereby a new distinction in their systems (cf. Table 13 and Section 9).

Finally, language contact can contribute to the development of temperature-term systems in various ways. We have seen occasional examples of temperature terms borrowed from one language into another (e.g., *proxladnyj* 'cool' in Russian from Church Slavonic, *zimny* 'cold' in Byelorussian from Polish, and *laden* 'cold' in Macedonian from Serbian). But on the whole borrowings seem to have been relatively marginal in the development of the modern Slavic temperature-term systems. On the other hand, language contacts may have been more instrumental in shaping the conceptual divisions in the systems themselves (as will be elaborated in Section 11 with respect to the distinction between 'warm' and 'hot').

---

27. The same contrast is also found between English ('cold' vs. 'hot'), on the one hand, and German and Swedish ('cold' vs. 'warm'), on the other hand, pointing to different placements of borders between 'warm' and 'hot'.

11. Slavic temperature-term systems in a broader cross-linguistic perspective

The temperature-term systems found across Slavic confirm several of the earlier (tentative) generalizations summarized in Koptjevskaja-Tamm (2015). First of all, the systems are very much in line with the main temperature-value systems found across the languages in Koptjevskaja-Tamm (ed) (2015), cf. Figure 2 in Section 1.

The distinctions between the frame-specific lexemes for a particular temperature value are also attested outside of Slavic, even though the details may vary. For instance, both Japanese and Hiw (Austronesian: Oceanic, spoken in northern Vanuatu) distinguish between tactile and non-tactile terms for 'cold' (François 2015) (cf. below for further examples).

Finally, the Slavic languages corroborate the earlier observations that temperature-term systems can show different degrees of temperature-value elaboration in their different subsystems defined in accordance to frames of temperature evaluation and/or entities. Ambient temperature evaluation often has a more elaborated system, and the same goes for temperature evaluation of water (cf. Koptjevskaja-Tamm 2015: 19–24, and Figure 1 in Section 1). As has been mentioned several times, in-between 'lukewarm' terms are normally restricted to water/liquids, with occasional additions of air and wind, which, again, is what is found in Slavic.

Cross-linguistically, the two-value termperature system appears to be the most widespread. It is found all over the world, including European languages such as Italian, Spanish and Portuguese, Greek, Armenian and Turkish. Macedonian is therefore in a good company, while being slightly exotic in the Slavic context.

The three-value and four-value systems found in Slavic are also found here and there among the world's other languages. For example, Indonesian *dingin* 'cold/cool' vs. *hangat* 'warm' vs. *panas* 'hot' (Siahaan 2015) and Finnish *kylmä* 'cold' vs. *lämmin* 'warm' vs. *kuuma* 'hot' (Juvonen & Nikunlassi 2015) exemplify three-value systems based on frame-neutral temperature terms. The three-value system in Yucatec Maya (Le Guen 2015) opposes the frame-neutral terms *chokoh* 'hot' and *k'iinal* 'warm' to the frame-specific terms *síis* 'cold/cool (tactile)' and *ke'el* 'cold/cool (non-tactile)'. Examples of four-value systems in which the distinctions are expressed by frame-neutral terms are the Baltic language Latvian (*auksts* 'cold', *vēss* 'cool', *silts* 'warm' and *karsts* 'hot', cf. Perkova 2015), the Sinitic language Cantonese (*dung[3]* 'cold', *loeng[4]* 'cool', *nyun[5]* 'warm' and *jit[6]* 'hot', cf. de Sousa et al. 2015), and the Araucanian language Mapudungan (*wütre* 'cold', *füshkü* 'cool', *eñum* 'warm', and *are* 'hot', cf. Zúñiga 2015). Elsewhere, the four-value systems may involve frame-specific temperature terms, as in the Sinitic language Southern Pinghua, which opposes the frame-neutral terms *jən[1]* 'cold', *lɛŋ[4]* 'cool' and *nun[5]* 'warm' to the frame-specific terms *lat[3]* 'hot (tactile)' and *ɲit[5]* 'hot (non-tactile') (de Sousa et al. 2015).

However, a word of caution is needed here. As has been shown in this paper, the status of 'cold' vs. 'cool' distinction varies across Slavic. In fact, it is generally unclear whether the 'cool' terms ever become as frequent as those for 'cold', 'warm' and 'hot' even when they may apply to all the three frames of temperature evaluation. There is therefore hardly any strict borderline between three- and four-value systems and different researchers may occasionally come to different conclusions on the centrality of terms in systems that are in some way comparable to each other.

But even if a distinction between 'warm' and 'hot' is attested in other languages, its stronghold seems to be Slavic, Baltic and Germanic within Indo-European, as well as Baltic Finnic and Hungarian (which have been in close contact with Slavic, Germanic and Baltic) within Finno-Ugric. This may therefore be a case of an areal lexico-semantic pattern, stemming from a combination of genetic and contact relations among the languages (Koptjevskaja-Tamm & Liljegren 2017).

Focusing more on the frames of temperature evaluation, the cross-linguistic norm seems to distinguish among them by different predicative constructions (cf. Pustet 2015 on the syntax of temperature predication), quite often involving different parts of speech. In this perspective, languages like English and Swedish, with their predicative construction ("Subject + copula + temperature adjective") being used for tactile, ambient and personal-feeling temperatures turn out to be relatively exotic. Slavic languages are quite interesting here in that the differences among the syntactic patterns are both persistent across the family and involve a regular alternation between adjectives and predicatives, rather than different words.

Finally, regarding stability of temperature terms, the Slavic temperature terms considered in this chapter are relatively stable and partly confirm earlier generalizations on lexical stability.[28] The discussion has shown that the issue of lexical stability is more complex than what is often assumed in large-scale investigations (e.g. Pagel et al. 2007). In addition, there seem to be significant differences among language families in the persistency of their temperature terms. For instance, the temperature terms in the two closely related Timor-Alor-Pantar languages Abui and Kamang (Schapper 2015), and across the Nyulnyulan family (Bowern & Kling 2015), are strikingly dissimilar. The meanings of cognates and their place in the overall temperature system of a language may be subject to significant variation.

A fascinating question is whether and/or to what extent different semantic domains differ in their degree of intra-genetic similarities vs. divergence. A first answer to this question is provided by Majid & Dunn's (2015) comparison of colour, body part, container and spatial relation domains in twelve Germanic languages. They demonstrate that there is more intra-genetic similarity in the semantics of

---

28. In Holman et al.'s (2008) stability ranking of the Swadesh 100-list 'cold' is ranked as 81 and 'hot' as 91.

containers than in spatial relations, whereas body parts and colours show most similarity overall. This chapter may serve as a model for close intra-genetic comparison of the temperature domain in other families and for the inclusion of this comparison in the broader comparison of different semantic domains.

Acknowledgements

Funding

Abbreviations

ACC	accusative	M	masculine
ADJ	adjective	N	neuter
AGR	agreement	NEG	negative
AOR	aorist	NOM	nominative
coll.	colloquial	PASS	passive
CVB	converb	PIE	Proto-Indo-European
COP	copula	PST	past
GEN	genitive	PL	plural
DAT	dative	PRED	predicative
DIAL	dialectal	PRS	present
F	feminine	PRF	perfective
GEN	genitive	PSl	Proto-Slavic
IMP	imperative	PTCP	participle
IPFV	imperfective	REFL	reflexive
INS	instrumental	SG	singular
LOC	locative	SUPERL	superlative

References

Ameka, F. 2015. "Hard sun, hot weather, skin pain": The cultural semantics of temperature expressions in Ewe and Likpe (West Africa). In Koptjevskaja-Tamm (ed), 43–72. https://doi.org/10.1075/tsl.107.02ame

Bedford, T. 1936. *Environmental warmth and its measurement*. Medical Research Memorandum 17. London: HMSO.

Bohnemeyer, J. & Brown, P. 2007. Standing divided: dispositionals and locative predications in two Mayan languages. A special issue of *Linguistics* 45(5): 1105–1151. https://doi.org/10.1515/LING.2007.033

Bowern, C. & King, L. 2015. Bardi temperature terms. In Koptjevskaja-Tamm (ed), 815–831. https://doi.org/10.1075/tsl.107.27bow

Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam/Philadephia: John Benjamins. https://doi.org/10.1075/slcs.71

François, A. 2015. Temperature terms in northern Vanuatu. In Koptjevskaja-Tamm (ed), 832–856. https://doi.org/10.1075/tsl.107.28fra

Goldberg, A. 1995. *A Construction-Grammar Approach to Argument Structure*. Chicago & London: University of Chicago Press.

Hensel, H. 1981. *Thermoreception and Temperature Regulation*. London: Academic Press.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A. & Bakker, D. 2008. Explorations in automated language classification. *Folia Linguistica* 42(2): 331–354. https://doi.org/10.1515/FLIN.2008.331

Juvonen, P. & Nikunlassi, A. 2015. Temperature adjectives in Finnish. In Koptjevskaja-Tamm (ed), 491–536. https://doi.org/10.1075/tsl.107.16juv

Koch, P. 2012. Location, existence, and possession: A constructional-typological exploration. *A special issue of Linguistics on New directions in lexical typology* 50(3): 533–604. https://doi.org/10.1515/ling-2012-0018

Koch, P. 2016. Meaning change and semantic shifts. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 21–66. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9783110377675-002

Koptjevskaja-Tamm, M. (ed). 2015. *The Linguistics of Temperature*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.107

Koptjevskaja-Tamm, M. 2015. Introducing "The linguistics of temperature". In M. Koptjevskaja-Tamm (ed), 1–40. https://doi.org/10.1075/tsl.107.01kop

Koptjevskaja-Tamm, M. & Liljegren, H. 2017. Lexical semantics and areal linguistics. In *The Cambridge Handbook of Areal Linguistics*, R. Hickey (ed), 204–236. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.009

Koptjevskaja-Tamm, M. & Rakhilina, E. 2006. "Some like it hot": on semantics of temperature adjectives in Russian and Swedish. *STUF (Sprachtypologie und Universalienforschung)* 59(2): 253–269. https://doi.org/10.1524/stuf.2006.59.3.253

Koptjevskaja-Tamm, M., Divjak, D. & Rakhilina, E. 2010. Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In *New Approaches to Slavic Verbs of Motion*, V. Driagina-Hasko & R. Perelmutter (eds), 315–341. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115.18kop

Kryvenko, A. 2015. In the warmth of the Ukrainian temperature domain. In M. Koptjevskaja-Tamm (ed), 300–332. https://doi.org/10.1075/tsl.107.10kry

Le Guen, O. 2015. Temperature terms and their meaning in Yucatec Maya (Mexico). In Koptjevskaja-Tamm (ed), 742–775. https://doi.org/10.1075/tsl.107.24gue

Letuchij, A. 2017. Predikativ [A predicative], <http://rusgram.ru/предикатив>.

Majid, A., Gullberg, M., van Staden, M. & Bowerman, M. 2007. How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics* 18(2): 179–194. https://doi.org/10.1515/COG.2007.007

Majid, A. & Dunn, M. 2015. Semantic systems in closely related languages. *Language Sciences* 49: 1–19. https://doi.org/10.1016/j.langsci.2014.11.002

Pagel, M., Atkinson, Q. D. & Meade, A. 2007. Frequency of Word-Use Predicts Rates of Lexical Evolution throughout Indo-European History. *Nature* 449 (7163): 717–720. https://doi.org/10.1038/nature06176

Perkova, N. 2015. Adjectives of temperature in Latvian. In Koptjevskaja-Tamm (ed), 216–253. https://doi.org/10.1075/tsl.107.08per

Plank, F. 2010. Temperature talk: The basics revisited. A talk presented at the Workshop on Temperature in Language and Cognition, Stockholm University, March 2010.

Pustet, R. 2015. The syntax of temperature predications. In Koptjevskaja-Tamm (ed), 889–916. https://doi.org/10.1075/tsl.107.30pus

Rakhilina, E. 2000. *Kognitivnyj analiz predmetnyx imën: semantika i sočetaemost'* [A Cognitive Analys of Entity Names: Semantics and Combinability]. Moscow: Russkie slovari.

Rakhilina, E. 2010. Verbs of rotation in Russian and Polish. In *New Approaches to Slavic Verbs of Motion*, V. Driagina-Hasko & R. Perelmutter (eds), 291–316. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115.17rak

Rakhilina, E. & Reznikova, T. 2016. A Frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–130. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9783110377675-004

Rasulic, K. 2015. What's hot and what's not in English and Serbian: A contrastive view on the polysemy of temperature adjectives. In M. Koptjevskaja-Tamm (ed), 254–299.

Schapper, A. 2015. Temperature terms in Kamang and Abui, two Papuan languages of Alor. In Koptjevskaja-Tamm (ed), 858–887. https://doi.org/10.1075/tsl.107.29sch

Siahaan, P. 2015. Why is it not cool? Temperature terms in Indonesian. In Koptjevskaja-Tamm (ed), 666–701. https://doi.org/10.1075/tsl.107.21sia

Sousa, H. de, Langella, F. & Enfield, N. J. 2015. Temperature terms in Lao, Southern Zhuang, Southern Pinghua and Cantonese. In Koptjevskaja-Tamm (ed), 594–638. https://doi.org/10.1075/tsl.107.19sou

Waldenfels, R. von. 2006. Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment. In *Beiträge der Europäischen Slavistischen Linguistik* (POLYSLAV) 9, B. Brehmer, V. Zdanova & R. Zimny (eds), 123–138. München: Otto Sagner.

Zimmerling, A. 2018. Predikativy i kategorija sostojanija v russkom jazyke [Predicatives and the category of state in Russian]. *Slavistična revija* 66(1): 45–64.

Zúñiga, F. 2015. Temperature terms in Mapudungun. In Koptjevskaja-Tamm (ed), 776–791. https://doi.org/10.1075/tsl.107.25zun

Dictionaries and other sources

Boryś, W. 2005. *Słownik etymologiczny języka polskiego* [An Etymological Dictionary of the Polish Language]. Kraków: Wydawnictwo literackie.

Černyx, P. Ja. 1994. *Istoriko-ètimologičeskij slovar' russkogo jazyka* [A Historical-Etymological Dictionary of the Russian Language], volumes 1–2. Moscow: Russkij jazyk.

Daľ, V. 1863–1866. *Tolkovyj slovarʹ živago velikorusskago jazyka* [An Explanatory Dictionary of the Living Great Russian language]. 1st ed. Moscow, Saint Petersburg, <http://slovardalja.net/>.

Digitalen rečnik na makedonskiot jazik [A Digital Dictionary of the Macedonian Language] <http://www.makedonski.info/>.

Dom. Leč. = B'juken, U., 1790–1792. *Polnyj i vseobščij domašnij lečebnik*. Tvorenie g. Buxana, Moskva, v. 1–5.

Dolnoserbsko-nimske słowniki [Lower Sorbian-German Dictionaries] <http://dolnoserbski.de>.

ÈSBM = *Ètymalahičny sloŭnik belaruskaj movy* [An Etymological Dictionary of the Byelorussian Language]. 1978–. Minsk: Navuka i tèxnika.

ESUM = *Etimologičnyj slovnyk ukraïns'koj movy* [An Etymological Dictionary of the Ukrainian Language], volumes 1–7. 1983. Kiev: Naukova dumka, <http://litopys.org.ua/djvu/etymolog_slovnyk.htm>.

Grinčenko, B. 1907–1909. *Slovar' ukraïns'koï movy* [A Dictionary of the Ukrainian Language], volumes 1–4. Kyiv, <http://hrinchenko.com>.

Hornjoserbsko-němski słownik [An Upper Sorbian-German Dictionary], <http://soblex.de>.

HSBM = *Gistaryčny sloŭnik belaruskaj movy* [A Historical Dictionary of the Byelorussian Language], volumes 1–21. 1982–. Minsk: Belaruskaja navuka.

Lom. SS = *Sobranie raznyx sočinenij v stixax i v proze… Mixajla Lomonosova*. 2-e izd. s pribavlenijami. 1757–1759. Moskva, 1–2.

Rejzek, J. 2001. *Český etymologický slovník* [A Czech Etymological Dictionary]. Prague: LEDA.

RUAS = *Rosijs'ko-ukraïns'kyj akademičnyj slovnik* [A Russian-Ukrainian Academy Dictionary], collected by A. Kryms'kij and S. Èfremov. 1924–1933. Kyiv: Červonyj šljax; Kharkiv: DVU, <http://r2u.org.ua/html/krym_details.html>.

Schuster-Šewc, H. 1980. *Historisch-etymologisches Wörterbuch der ober- und niedersorbischen Sprache*. Bautzen: Domowina-Verlag.

Skarnik. *Russko-belorusskij slovar'. Belorussko–russkij slovar'. Tlumačaľny sloŭnik belaruskaj movy* [A Russian-Byelorussian dictionary. A Byelorussian dictionary. An Explanatory Dictionary of the Byelorussian Language], <http://www.skarnik.by/>.

Skok, P. 1971–1974. *Etimologijski Rjecnik Hrvatskoga Ili Srpskoga Jezika* [An Etymological Dictionary of the Serbian or Croatian language], 4 volumes. Zagreb: Jugoslavenska Akademija Znanosti i Umjetnosti.

Sl. Komm. = *Slovar' kommerčeskij, soderžaščij poznanie o tovarax vsex stran, i nazvanijax veščej glavnyx i novejšix, otnosjaščixsja do kommercii*. Perevod s francuzskogo jazyka Vasil'em Levšinym. Moskva, 1787–1792, 1–7.

Slovar slovenskega knjižnega jezika [Dictionary of Standard Slovene]. 1970. (1a ed. and numerous later editions). Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, <http://www.fran.si/201/sskj3-slovar-slovenskega-knjiznega-jezika>.

Snoj, M. 2016. Slovenski etimološki slovar [A Slovene Etymological Dictionary]. Ljubljana: Založba ZRC, ZRC SAZU, <http://www.fran.si/193/marko-snoj-slovenski-etimoloski-slovar>.

Spb. Ž. = *S. Peterburgskij žurnal izd. I Pninym*. 1798 SPb., [1798], 1–4.

SPI = *Slovar'-spravočnik "Slova o polku Igoreve"* [A Dictionary–Handbook of "The Tale of Igor's Campaign"], 1965–1984, volumes 1–6. Leningrad: Nauka, <http://feb-web.ru/feb/slovoss/ss-abc/>.

Sreznevskij, I. I. 1987. *Slovar' drevnerusskogo jazyka* [A Dictionary of the Old Russian Language], volumes 1–3. Reprinted edition. Moscow: Kniga.

SRJa XI–XVII = *Slovar' russkogo jazyka XI–XVII vekov* [A Dictionary of the Russian Language of the 11th – 17th Centuries]. 1975–. Moscow: Nauka.

SRJa XVIII = *Slovar' russkogo jazyka XVIII veka* [A Dictionary of the 18th Century's Russian]. 1984–, volumes 1–21. Leningrad (until 1991) /Saint Petersburg: Nauka, <http://feb-web.ru/feb/sl18/slov-abc/>.

SUM = *Slovnyk ukrains'koj movy. Akademičnyj tlumačnyj slovnyk 1970–1980* [A Dictionary of the Ukrainian Language. An Academy Explanatory Dictionary], volumes 1–11. Kyiv: Naukova dumka, <http://sum.in.ua>.

Lexical typology of Mandarin Chinese qualitative features

Liliya Kholkina
Russian State University for the Humanities

The chapter analyses the quality terms of Mandarin Chinese which belong to the semantic domains of ʜᴇᴀᴠʏ and ʜᴀʀᴅ, as well as to the subdomains of surface texture (ꜱʟɪᴘᴘᴇʀʏ, ꜱᴍᴏᴏᴛʜ, and ʟᴇᴠᴇʟ). We argue that the evidence from Mandarin Chinese accords with the larger typological picture which has been gleaned from languages that are areally and genetically distant from Chinese. This similarity is firstly observed in the literal meanings, which are discussed below for the subdomains of ꜱʟɪᴘᴘᴇʀʏ, ꜱᴍᴏᴏᴛʜ, and ʟᴇᴠᴇʟ. Furthermore, the Mandarin data supports the existence of stable links between literal and figurative meanings. With data from the ʜᴀʀᴅ domain, we show that the oppositions between the lexemes' literal meanings determine the results of their semantic shifts. Next, we examine evidence from the ʜᴇᴀᴠʏ domain to demonstrate that a reverse analysis can be applied to predict differences in literal meanings from the differences in metaphoric uses. The concluding sections of the chapter discuss how lexico-typological data can contribute to Chinese lexicology in general.

Keywords: Mandarin Chinese, lexicography, heavy, hard, smooth, synonyms

Introduction

Due to a number of reasons, Mandarin Chinese presents a case of special interest for lexico-typological studies. Firstly, the traditional Chinese culture, its system of values, and the way of life are dramatically different from those of European societies. If one assumes that these factors are likely to have a profound effect on the system of lexical oppositions in a language, then Chinese could be expected to reveal non-trivial evidence that would not match with the typological generalizations presented in Chapters 1 and 6 of this volume (see also Ryzhova et al. 2019). However, this chapter defies such expectations by showing that the Mandarin Chinese lexical system is generally comparable to the systems found in European languages and, therefore, they can be treated on equal terms in lexico-typological studies.

Secondly, Mandarin Chinese is remarkably abundant in rich sets of synonyms. One of the reasons (and, in case of Mandarin Chinese, probably, the major one), is its continuous written tradition going back thousands of years: many idioms and means of expression that can be traced back to Old Chinese are retained in the modern language along with their innovative correlates. Scattered fragments of the bygone stages of Chinese are preserved in modern Sinitic languages/dialects, with different varieties reflecting the lexicons of different periods (Jiang 2015). In addition, modern Chinese mass media lavishly introduce dialectal vocabulary into the national standard; on the one hand, this practice leads to enrichment of the vocabulary, but on the other, it further complicates the task of describing relevant lexical fields.

The third reason contributing to the importance of Mandarin Chinese for lexical typology lies in its another well-known peculiarity – namely, the abundance of lexemes that are formed via root compounding. This feature also contributes to synonymy, due to the existence of multiple compounds that share roots and engage in synonymic relations of the following kinds: (1) one word in a pair of synonyms can be monosyllabic and the other disyllabic, e. g. 尖 *jiān* 'sharp, pointed' and 尖锐 *jiānruì* (*jiān* 'sharp, pointed' + *ruì* 'keen') 'sharp-pointed, keen'; (2) both words can be disyllabic and share a single root between them, e.g. 尖锐 *jiānruì* 'sharp-pointed, keen' and 尖利 *jiānlì* (*jiān* 'sharp, pointed' + *lì* 'sharp') 'sharp; keen; cutting'; or even (3) two disyllabic words can be made up of the same pair of roots but these roots are arranged in a different order, e.g. 空虚 *kōngxū* (*kōng* 'empty' + *xū* 'void') 'empty, hollow, void' and 虚空 *xūkōng* 'hollow, void', etc. (Semenas 1992; Pan et al. 1993). Relations between the meanings of monosyllabic and disyllabic synonyms constitute a separate problem which lies outside the scope of this chapter.[1] However, we believe that this aspect of Chinese word-formation suggests a promising and compelling direction for future research in lexical typology.

By virtue of the above factors, lexico-typological description of Mandarin Chinese is deemed to constitute an interesting and challenging task; its results are expected to make a valuable contribution to the previous typological research, including the findings of the Moscow Lexical Typology group (MLexT).

The approach proposed by the MLexT group (Rakhilina & Reznikova 2016; Chapter 1, this volume), which is also adopted in the present study, claims that each semantic domain corresponds to a set of frames. Frames are understood as typical situations that may be expressed by specialized lexemes in some languages.

---

1.   The use of monosyllabic and disyllabic tokens is also subject to stylistic and rhythmic restrictions (Ke 2012). Thus, colloquial speech is characterized by predominant use of monosyllabic tokens, while written language tends to use disyllables. As for poetic language, it also encourages use of monosyllables, a feature inherited from Classical Chinese with its largely monosyllabic lexicon (Feng 2009, 2015, 2017).

For instance, the semantic domain SHARP comprises three different situations, or frames: sharpness of 'instruments with a functional edge' (cf. *sharp knife*), sharpness of 'instruments with a functional end-point' (cf. *sharp spear*), and sharpness of 'objects of a pointed shape' (cf. *sharp nose*). Frames are identified when the boundaries between them coincide with the lexical opposition within the domain (for more detail, see Chapter 2 of this volume). At the same time, there are cases when different frames may be covered with only one lexeme (or 'colexified', see François 2008: 170); thus, the number of lexemes in a given domain can vary from language to language. Different strategies of colexification are visualised by means of semantic maps, where frames are depicted as nodes (the architecture of semantic maps is analogous to the design which is accepted in grammatical typology, see Haspelmath 2003, cf. also Georgakopoulos & Polis 2018; Rakhilina et al., in press). Semantic maps prove to be instrumental in making predictions about the viable as well as the impossible strategies of (co)lexification in a given semantic domain. Previous research has shown that figurative meanings evolve along similar lines in various languages, and their evolution is conditioned by the oppositions that are present in literal meanings. Consequently, metaphoric uses can be predicted and described by the methods of lexical typology.

This chapter will examine Mandarin Chinese data with respect to two types of typological predictions. On the one hand, we will focus on the structure of several semantic domains in Chinese and how they comply with typological generalizations made on the basis of other languages. On the other hand, we will analyse not only the literal but also the figurative meanings developed by the lexemes under consideration, since the richness and complexity of Mandarin Chinese data hold promise of extraordinary and surprising findings in the sphere of metaphorical uses.

The chapter is divided into four sections. Section 1 looks at the literal meanings of qualitative lexical items; we will discuss the semantic domain which is already familiar to the reader (Chapter 6, this volume), ABSENCE OF IRREGULARITIES ON THE SURFACE. We will see to what extent the Mandarin Chinese evidence conforms to the larger typological picture which has been reconstructed on the basis of languages that are areally and genetically distant from Chinese. Sections 2 and 3 concentrate on figurative meanings of qualitative lexemes. We will first address the domain HARD by analysing the instances that conform to our typological expectations. This will allow us to test the hypothesised existence of stable links between literal and figurative meanings against novel and non-trivial material. Further, we will concentrate on the domain HEAVY and discuss the case where Mandarin Chinese exhibits a somewhat typologically divergent behaviour. We will show that this evidence offers a new perspective on the data of the previously studied languages. Finally, Section 4 will show how lexico-typological data can contribute to Chinese lexicology in general.

1.   ABSENCE OF IRREGULARITIES ON THE SURFACE: Typological expectations

This section addresses the task of assessing the degree to which Mandarin Chinese literal qualitative meanings are congruent with the expectations that follow from prior typological research. For the sake of convenience, let us begin with a discussion of the semantic domain that has been already discussed in this book (Chapter 6), ABSENCE OF IRREGULARITIES ON THE SURFACE.

The primary opposition within this domain is found between the tactile and the visual modes of perception. In the first mode, the properties of a surface are evaluated by tactile contact with it ('*smooth* wooden board', '*slippery* road'), and in the second – via general, typically visual, assessment (for example, '*flat* fields', '*level* ground'); the latter mode belongs to the LEVEL subdomain of the domain under examination, which will be discussed in the end of this section. As for the frames of tactile perception, they can belong either to the SMOOTH subdomain (for example, 'smooth shaven cheeks' or 'smooth book cover') or to the SLIPPERY subdomain ('slippery icy road' or 'fish slipping out of one's hands'). These two subdomains have the same idea in common, that of low friction between the contacting surfaces; in the case of SLIPPERY, the friction is not sufficient for an object to be held in one's hands or to stay reliably on a certain surface.

The domain ABSENCE OF IRREGULARITIES ON THE SURFACE appeared to be very rich in Mandarin Chinese; in other languages it is covered by four lexemes, on average (Chapter 6, this volume), while in Chinese there are nine. Three monosyllabic tokens – 滑 *huá* 'slippery, smooth', 光 *guāng* 'smooth, shiny', and 平 *píng* 'level' – correspond to the three main subdomains that have been identified in this domain typologically (SLIPPERY, SMOOTH, and LEVEL); in addition, there are six disyllables associated with these monosyllables: 滑溜 *huáliu* 'slippery, smooth', 平滑 *pínghuá* 'smooth, flat', 光滑 *guānghuá* 'smooth, sleek', 光溜 *guāngliu* 'smooth, glossy', 平坦 *píngtǎn* 'level, even, plane', and 平整 *píngzhěng* 'levelled, even'. The interesting questions that arise from this are: How does such a prolific system correlate with our typological predictions? How can so many lexemes be arranged on the semantic map? Does this system give rise to new oppositions? To answer these questions, we will take a closer look at all of the above-mentioned Mandarin Chinese lexemes.

To begin with, let us consider the monosyllabic tokens, as they convey simpler, basic notions. The primary meaning of 滑 *huá* 'slippery, smooth' is the low friction that occurs when one touches objects or surfaces. For such frames as 'slippery surface', 'slippery sole', and 'slippery object', 滑 *huá* 'slippery' and its disyllabic synonym 滑溜 *huáliu* 'slippery, smooth' (where the meanings of both syllables are identical, i.e. 溜 *liu* literally also means 'slippery, smooth') are the only possible lexemes. Therefore, both 滑 *huá* and 滑溜 *huáliu* are presumed to have 'slippery' as their primary meaning. At the same time, these two lexemes can occasionally describe

smooth objects and surfaces as well as 'smooth human skin', e.g. 她的皮肤很滑 *Tā de pífū hěn huá* 'Her skin is very smooth' (see Figure 1; for the semantic map of the whole domain, see Chapter 6 of this volume).
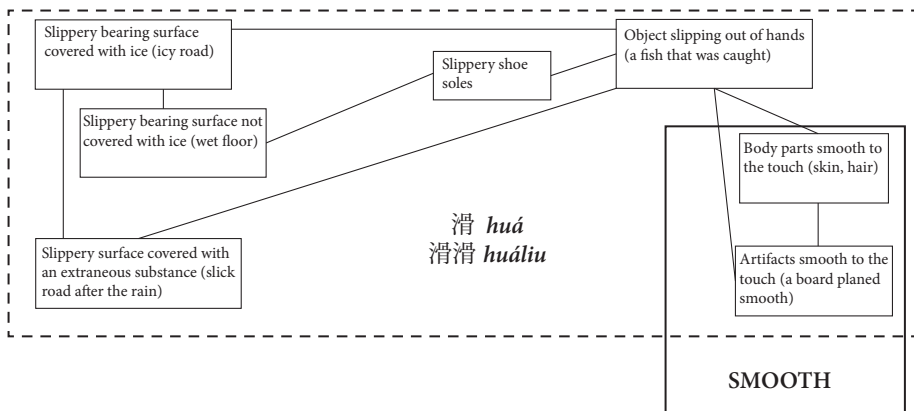
Figure 1. ABSENCE OF IRREGULARITIES: SLIPPERY (Mandarin Chinese)

The replacement of 滑 *huá* 'slippery, smooth' by 滑溜 *huáliu* 'slippery, smooth' is possible in almost all contexts both in the SLIPPERY and the SMOOTH subdomains, except for compounds and lexicalized collocations. The difference between these two lexemes lies not only in rhythmic and stylistic limitations but also in the additional positive evaluation which is often carried by 滑溜 *huáliu* in the sense of 'smooth'. This additional positive hue of 滑溜 *huáliu* is pointed out in Chinese dictionaries (XHC 6. 2015).

The colexification of the SLIPPERY and SMOOTH subdomains, which is characteristic of both 滑 *huá* and 滑溜 *huáliu*, has typological parallels. For example, in Khanty (the dialect of Tegi), the adjective *woľʼǝk* (with the allomorph *woľʼk*) means both 'slippery' and 'smooth', whereas 'level' is denoted by another adjective, *pajłi* (Chapter 6, this volume).

Another disyllable that contains the component 滑 *huá* is 平滑 *pínghuá* 'smooth, flat' (lit. 'level-smooth'). This disyllable inherits only one of the two meanings present in 滑 *huá* 'slippery, smooth', namely, 'smooth'. Furthermore, 平滑 *pínghuá* demonstrates some selectivity even with smooth objects, excluding nonplanar surfaces (for example, *平滑的石球 *pínghuá de shíqiú* 'a smooth stone ball'). Despite this, 平滑 *pínghuá* can be referred to the SMOOTH subdomain.

Among the monosyllables, the main lexeme for 'smooth' is 光 *guāng* 'smooth, shiny'. However, it is used quite rarely in this meaning (which is currently of interest to us). As a consequence, even though the dictionary (XHC 6. 2015) illustrates this meaning with the sentence 这种纸很光 *zhè zhǒng zhǐ hěn guāng* 'This paper is very

smooth', many informants have suggested replacing 光 *guāng* with its disyllabic counterpart, 光滑 *guānghuá* 'smooth, sleek' so as to convey the meaning of the sentence more accurately. According to these elicitations from our informants, it can be claimed with certainty that in the synchronic perspective, the monosyllable 光 *guāng* is only used as a resultative or an evaluative morpheme in the sense of 'smooth' (for example, 磨光 *móguāng* ('to polish' + 'smooth') 'to polish smth. smooth'), although historically the scope of its use was wider (cf. the Classical Chinese examples where 光 *guāng* describes hair (1) and mirror (2)).

(1) 梁　世　　有　人，　常　以　雞卵白　　　和　沐，　云
     Liáng shì　yǒu rén　cháng yǐ jīluǎn bái　huò mù　yún
     Liang period have person often use white of the egg mix wash_hair say
     使　髮　光。
     *shǐ　fà　**guāng***
     make hair smooth_and_shiny
     'There was a man during the Liang period who often washed his hair with egg whites, as he believed it would make his hair smooth and glossy.'

《顏氏家訓/歸心第十六》6 C.

(2) 如　一　镜　然，　　今日　磨　些，　明日　磨　些，　不
     rú　yī　jìng　rán　jīnrì mó　xiē　míng rì　mó　xiē　bù
     like one mirror that_way today polish some tomorrow polish some NEG
     觉　自　光。
     *jué　zì　**guāng***
     sense (it)self smooth_and_shiny
     'It is like a mirror that you polish a little today, then a little tomorrow, so that you will not even notice how it becomes glossy.'　　《朱子语类·卷五》13 C.

We can see that in these examples the features of smoothness and glossiness are associated with each other. However, from the historical point of view, the meaning 'shiny, glossy' is an earlier one; in *Shuowen Jiezi*, an etymological dictionary of Chinese hieroglyphs compiled in 2nd century CE, the hieroglyph 光 *guāng* is listed under the classifier 'fire'. In the modern language, according to the *Dictionary of Modern Chinese*, the most frequent meaning of 光 *guāng* is that of 'light'.

The connection between smoothness and glossiness is attested in other languages, cf. the English adjective *sleek* meaning 'smooth and shining' vs. the more neutral term *smooth* (see Chapter 6, this volume); however, in Chinese the interrelation between these two concepts can also be seen at the level of word polysemy.

The most frequent lexeme for conveying the meaning of smoothness has already been mentioned above – the disyllabic 光滑 *guānghuá* 'smooth, sleek'. This disyllable has a wide degree of compatibility: it can be used to describe carrying surfaces (road), surfaces where potential irregularities are normally invisible and can

be perceived only tactilely (the marble surface of a table, the cover of a magazine), as well as artefacts and objects that can be perceived both tactilely and visually, e.g. human skin and hair. In the latter case, 光滑 *guānghuá* often preserves the inherent idea of the glossy characteristic of 光 *guāng*, see (3):

(3) 淡淡　　的　斜阳，　　照着　　　她　丝绸　般　柔软　　光滑
　　 *dàndàn de xiéyáng   zhàozhe   tā sīchóu bān róuruǎn **guānghuá***
　　 Weak   ATR setting_sun reflect.PRG she silk like soft   smooth
　　 的　头发
　　 de   tóufa
　　 ATR hair
　　 'The weak rays of the setting sun reflected themselves in her hair, soft and smooth like silk.'

The second lexeme pertaining to the domain SMOOTH is the disyllable 光溜 *guāngliu* 'smooth, glossy'. The literal meanings of its two roots are 'smooth, shiny' and 'slippery, smooth', respectively. The *Dictionary of Modern Chinese* defines 光溜 *guāngliu* in the same way as 滑 *huá* 'slippery, smooth'; yet, according to our data, this definition is erroneous. For 滑 *huá*, the meaning 'slippery' is very important because (along with the disyllable 滑溜 *huáliu* 'slippery, smooth') it is the only way to convey this meaning in Mandarin Chinese. By contrast, 光溜 *guāngliu* does not place any emphasis on the fact that a surface is difficult to keep one's balance on, or that an object is difficult to be held in one's hands; consequently, the notion of smoothness comes to the fore. However, this interpretation raises the question of the difference between 光溜 *guāngliu* 'smooth, glossy' and the main lexeme for 'smooth', 光滑 *guānghuá* 'smooth, sleek'.

To distinguish between the semantics of 光滑 *guānghuá* and 光溜 *guāngliu*, we resorted to questionnaires and asked the informants to fill out the blanks in the sentences with one of the suggested words.[2] In the majority of the test sentences, the respondents gave preference to 光滑 *guānghuá* (which is a much more frequently used word, generally); however, in a small number of contexts – such as 'smoothly combed hair',[3] 'bald head', or 'clean-shaven chin'– 90% of the choices were made in favour of 光溜 *guāngliu*.

---

2.   Responses were collected from 32 students of different specializations from Peking University; the respondents represented a variety of dialectal backgrounds.

3.   As we noted above, 光滑 *guānghuá* can also describe hair, but it applies to naturally smooth hair, while 光溜 *guāngliu* foregrounds the idea of hair that is smoothly combed (so that no hairs are entangled or stick out).

(4) 他的　禿　头　像　　台球　　　一样　光溜。
　　 *tā de　tū　tóu　xiàng　táiqiú　　yíyàng **guāngliu***
　　 he ATR bald head look_like billiard_ball alike   smooth
　　 'His bald head is smooth, just like a billiards ball.'

According to the comments provided by the respondents, they use 光溜 *guāngliu* 'smooth, glossy' when they mean to place emphasis on the lack of hairs or fibers or other protruding objects on a surface (such as on a splintery wooden board).[4] Thus, Mandarin data allows us to postulate an additional parameter for the subdomain SMOOTH: it opposes 'absence/presence of individual projections on the surface" to other types of smooth surfaces. This is a crucial observation, and the newly attested parameter can be regarded as a contribution of the Mandarin Chinese language to the typology of lexemes describing surface texture. This additional parameter cannot be reflected in any way on the semantic map suggested below (Figure 2); nevertheless, it is evident that the types of objects that can be described with 光溜 *guāngliu* 'smooth, glossy' tend to group together. In particular, each lexeme covers only nodes that are directly connected with each other; the mapping thus adheres to what is called "the principle of contiguity" (Haspelmath 2003).
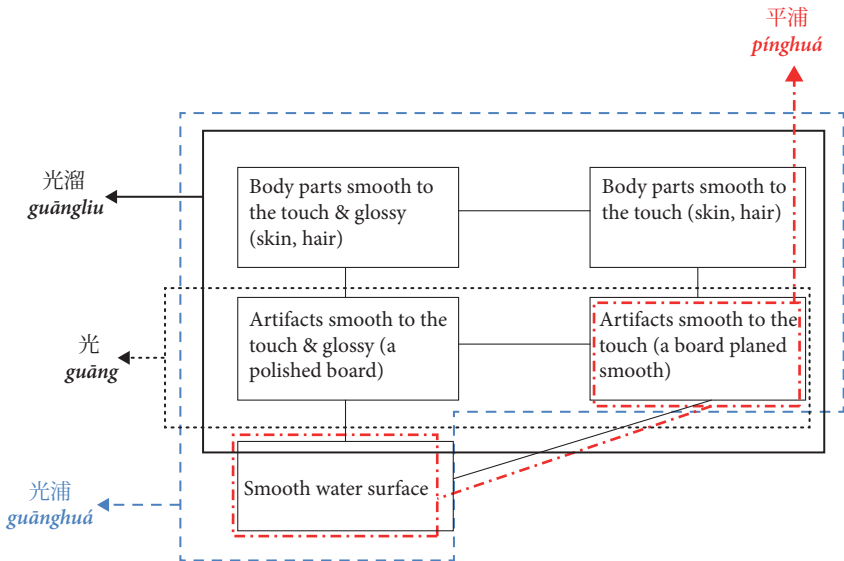
Figure 2. ABSENCE OF IRREGULARITIES: SMOOTH (Mandarin Chinese)

---

4.   The respondents also indicated that an additional parameter influencing the lexical choice is the preferably spherical shape of the object described by 光溜 *guāngliu*. However, this parameter is not as relevant as the other parameters described above, since 光溜 *guāngliu* was reported to be applicable to flat objects as well.

Let us now move on to the subdomain LEVEL. The central word for describing the visual perception of levelness is 平 *píng* 'level', and its use as a monosyllable is relatively frequent. The further refinement of this meaning is accomplished with the use of the disyllabic tokens 平坦 *píngtǎn* 'level, even, plane' and 平整 *píngzhěng* 'level, even'.

The disyllable 平坦 *píngtǎn* 'level, even, plane' (lit. 'even-wide') is used mainly to describe landscapes; it most often applies to the names of such objects as plains, shallows, pastures, meadows, or land plots, but it can also be used with man-made objects such as roads.

As an adjective, the word 平整 *píngzhěng* 'level, even, plane' (lit. 'even-ordered') describes a surface that has become flat after some effort has been exerted on it. The most frequent are the combinations of 平整 *píngzhěng* with names of various artefacts that have a flat surface and can be oriented both horizontally (floor, road) and vertically (wall, fence) – see Figure 3.
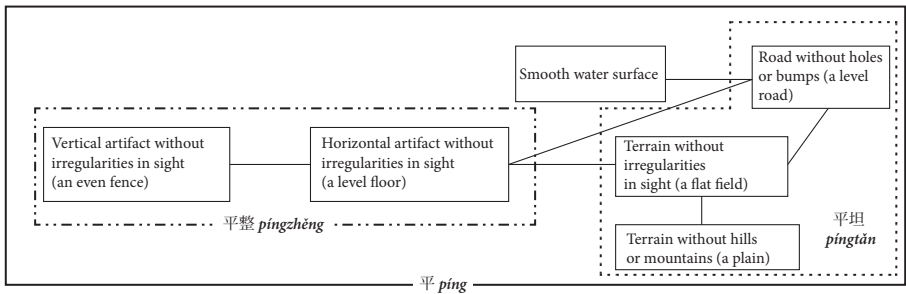
Figure 3. ABSENCE OF IRREGULARITIES: LEVEL (Mandarin Chinese)

Figures 1–3 demonstrate that the three basic subdomains comprising the ABSENCE OF IRREGULARITIES domain are covered in Mandarin Chinese by three different monosyllables (滑 *huá*, 光 *guāng*, and 平 *píng*), all the finer oppositions being specified with disyllables. The fact that the most basic oppositions are covered by monosyllables, and the more subtle distinctions within the subdomains rely on disyllabic units, corroborates the fundamental principle of semantic organization of the Chinese lexicon: "In Chinese, disyllables are very rarely used to express categories of a basic level" (Ye 2004: 100).

To summarize, the lexical richness and uniqueness of Mandarin Chinese has allowed us to highlight interesting semantic nuances that can prove significant for the domain under consideration, such as the "absence of individual projections on the surface" (光溜 *guāngliu*) in the subdomain of SMOOTH. It should be noted that we have not attested any cases of non-trivial combinations of meanings in one lexeme, as all of the examined lexemes can be plotted on semantic maps without

violating the principle of contiguity. In general, the Mandarin Chinese semantic maps showed no significant differences from other languages. Thus, we can conclude that when it comes to literal meanings, the Chinese data largely conforms to typological predictions.

2.   HARD: Literal and figurative meanings

Let us now turn to figurative meanings, this time using a different domain for illustration. The vast literature on metaphor and the practice of its analysis have shown that metaphorical shifts are not accidental; instead, they follow certain recurrent patterns (see, e.g., Lakoff & Johnson 1980; Lakoff 1987; Apresjan 2000; Zalizniak et al. 2012). The studies conducted within the MLexT framework advanced a much more precise version of this observation. It has been demonstrated that it is not the literal meaning of a word that should be regarded as the starting point of the metaphorical shift; rather, the shift is motivated by the frames expressed by a given word.

Let us illustrate this with the semantic field SHARP (see Chapter 2 of this volume). The functional usages of this semantic field embrace two typologically relevant frames, i.e. 'sharp instruments with a functional edge' (*sharp knife, sharp sword*) and 'sharp instruments with a functional end-point' (*sharp needle, sharp spear*).[5] While some lexemes with the meaning 'sharp' in languages can describe both types of objects (English *sharp*, Russian *ostryj*, Finnish *terävä*, Malay *tajam*), others are confined to only one frame. For example, the French lexeme *tranchant* describes objects with a cutting edge, but not with a piercing point (*un couteau tranchant* 'a sharp knife', but **une aiguille tranchante* 'a sharp needle'). The same is true for the German word *scharf*, the Komi-Zyrian (Uralic) *lečyd*, the Kla-dan (Mande) *líɛ́ɛ̀e*, among others. The frame 'sharp instruments with a functional end-point' in these languages is covered by a different lexeme, i.e. *aigu* (French), *spitz* (German), *jues'* (Komi), *zűɛ́ɛ̀* (Kla-dan).

The analysis of figurative meanings developed by lexemes that cover only one frame in their literal usages shows that, at least, some metaphors can only occur if a lexeme covers a specific frame. For example, the meaning 'clear line / with clear lines' can be only developed by a lexeme of the field SHARP if it covers the frame 'sharp instruments with a functional edge' in its direct usages, compare English phrases *razor-sharp knife* → *razor-sharp image*. On the contrary, the metaphor 'penetrating look' is derived from the frame 'sharp instruments with a functional end-point', see French *un javelot aigu* 'a sharp spear' → *un regard aigu* 'a keen look'.

---

5.   The set of the core frames of this field also includes a non-functional situation 'objects of a pointed shape' (toe, nose), but we are not going to include this frame in the current discussion.

We postulate that the revealed connections between the frames and metaphors are unbreakable, i.e. there will be no lexeme in any language that would only cover the frame 'sharp instruments with a functional edge', for example, and will develop the metaphor 'penetrating look' at the same time. In other words, the opposition between frames in the literal meanings is reflected in the figurative meanings developed by a corresponding lexeme (see Chapter 1 of this volume, cf. also Rakhilina & Reznikova 2016). The figurative meanings of a lexeme which collapses two or more literal frames also conflate the metaphors that arise from the colexified frames. Cf. English *sharp* that colexifies properties of an 'instrument with a cutting edge' (*sharp knife*) and 'instrument with a functional end-point' (*sharp spear*) and develops metaphors of both types: *sharp wind* (*image*), *sharp mind*.

The present section addresses the above regularity in Mandarin Chinese using the example of the domain HARD, which has been investigated within the framework of MLexT in the broader typological perspective (see Pavlova 2014; Pavlova & Kholkina 2013).

According to previous research, an important opposition within the domain HARD is whether a given lexeme describes the property per se or its influence on the experiencer. For example, the Russian adjective *tverdyj* 'hard' describes the feature of the object itself, irrespective of any evaluative connotations, as in (5)–(6); while *žestkij* 'hard' describes an object that not only resists deformation but also causes inconvenience to the experiencer, cf. (7)–(8).

(5) *Drevesina akacii,   bolee temnaja   i   **tverdaja**,   čem*
     wood:NOM acacia:GEN.SG more dark:NOM.F.SG and hard:NOM.F.SG than
     drevesina duba,   črezvyčajno dolgovečna ⟨…⟩.   [RNC][6]
     wood:NOM oak:GEN.SG extremely   durable:F.SG
     'Acacia wood, which is darker and harder than oak wood, is extremely durable.'

(6) *Sovet: rabotajte   na **tverdoj**   poverxnosti,   skladyvajte*
     tip:NOM.SG work:IMP.PL on hard:LOC.F.SG surface.LOC.SG fold:IMP.PL
     bumagu   s   pomošč'ju linejki ⟨…⟩   [RNC]
     paper:ACC.SG with help:INS   ruler:GEN.SG
     'Here's a tip: work on a hard surface, fold the paper with a ruler.'

(7) ⟨…⟩ *ona   nasil'no ževala   kusočki   **žestkogo**   mjasa.*
             she:NOM by_force chew:PST.F.SG piece:ACC.PL hard.GEN.N.SG meat.GEN
     'She forced herself to chew on the pieces of hard meat.'   [RNC]

(8) ⟨…⟩ *u menja neudobnaja   **žestkaja**   krovat'* ⟨…⟩
             at I:GEN uncomfortable:NOM.F.SG hard:NOM.F.SG bed:NOM.SG
     'I have a hard, uncomfortable bed.'   [RNC]

---

6. Russian National Corpus (available at www.ruscorpora.ru, accessed on 2020-11-06).

This opposition is retained in the figurative meanings. Metaphorically, *tverdyj* denotes an abstract characteristic of a certain phenomenon as not susceptible to change (*tverdye ceny* 'fixed prices', *tverdyj kurs* 'fixed rate') or not susceptible to outside influences (*tverdyj čelovek* 'a strong-willed person', *tverdyj xarakter* 'firm character', *tverdoe rešenie* 'steadfast decision'). In its figurative uses, *žestkij* still accentuates the experiencer's perception and, possibly, an unpleasant experience of interacting with some unyielding power (*žestkie mery* 'drastic measures', *žestkij kontrol′* 'tight control', *žestkij xarakter* 'harsh temper').

The primary lexeme of the domain HARD in Mandarin is 硬 *yìng*. It can be used to describe the physical property of being hard, as in (9), as well as an unpleasant interaction with a hard object, as in (10).

(9) 注入 碳　氮　　粒子 后　产生　　 了 金属 的　碳化物　或
　　 zhùrù tàn,　dàn　lízǐ　hòu, chǎnshēng le　jīnshǔ de　tànhuàwù huò
　　 inject carbon nitrogen ion　 after produce　PFV metal ATR carbide　 or
　　 氮化物，　这　些　　化合物　　很　硬
　　 dànhuàwù zhè xiē　huàhéwù　hěn yìng
　　 nitride　　 this some compound very hard
　　 'After the introduction of carbon and nitrogen ions, carbides or metal nitrides
　　 were obtained, these compounds being very hard.'　　　　　　　　 [ccl]

(10) 花蛤 肉　粗　而　硬，咬　不　动。
　　 huāgé ròu　cū　ér　yìng yǎo bù　dòng
　　 clam　meat raw and hard bite NEG move
　　 'Clam meat is raw and tough, it is hard to chew.'　　　　　　　 [ccl]

The figurative uses of 硬 *yìng* also develop in two directions. In Examples (11)–(12) this lexeme describes an impersonal and rather positive characteristic of an object. Here, 硬 *yìng* describes a reliable product of high quality that can endure long-lasting and intensive usage (11), as well as fixed rate which is capable of withstanding fluctuations under external influence in (12). Conversely, in (13)–(14) 硬 *yìng* refers to a discomforting interaction with an object, emphasizing the experiencer's perception of it, cf. the stringent requirements that place pressure on those who are subject to them (13), or the harsh tone of one's voice, which produces an unpleasant impression on the hearer (14).

(11) 这　东西　是 硬　货。
　　 zhè dōngxi shì yìng huò
　　 this thing　be hard product
　　 'This thing is a product of good quality.'

(12) 兑换码　　　很　硬。
duìhuànmǎ　*hěn*　**yìng**
exchange_rate very hard
'The exchange rate is very stable.'

(13) 为 此，市 里 给 各　乡镇　　　　　下达　 硬　指标
wèi cǐ　*shì lǐ gěi gè*　*xiāngzhèn*　　　*xiàdá*　**yìng** *zhǐbiāo*
for this city in for every villages_and_towns transmit hard index target
完　不　成　　　　任务，要　受到　　处罚。
wán　*bu*　*chéng*　　*rènwu yào shòudào chǔfá*
finish NEG accomplish task　must receive　punishment
'To this end, strict requirements were imposed by the city on all villages and small towns; if the task is not fulfilled, punishment must be administered.'

(14) 两　个 头目 看见 张鼐　　的　脸色　　　　　严峻，问话
liǎng gè　*tóumù kànjiàn Zhāng Nài de*　*liǎnsè*　　　　*yánjùn wèn huà*
two CLF chief see　Zhang Nai ATR facial_expression strict　question
的　口气　　很　硬，感到　不妙。
de　*kǒuqì*　*hěn*　**yìng** *gǎndào bú miào*
ATR intonation very hard feel　　NEG fine
'The two leaders saw Zhang Nai's face grow stern, and the intonation of his question was harsh, so they sensed that the situation was not alright.'

These examples support the claim that the figurative meanings of 硬 *yìng* retain the insensitivity to the opposition between an object's feature per se and the impact of the feature on the experiencer, which we established in the literal meanings.

This finding leads to the assumption that differences in the figurative meanings may be indicative of differences in the literal meanings of lexemes. In order to illustrate and test this conjecture, let us consider the domain HEAVY in Section 3 below.

3. HEAVY: Diagnosing oppositions within literal meanings through figurative ones

The domain HEAVY has been examined along with several other domains within the project on qualitative features carried out by the MLexT group (Kyuseva et al. 2012; Ryzhova et al. 2019). As long as the project concentrated on data from European languages, only negative connotations were found in the figurative meanings of the lexemes of the domain HEAVY, cf. French *pensées lourdes* 'heavy thoughts', *lourd destin* 'hard destiny', German *schwerer Eindruck* 'painful impression', and Russian *tjaželyj fil'm* 'disturbing film', *tjaželaja zadača* 'hard task', etc.

However, in Mandarin, such negative connotations in figurative meanings turned out to be uncharacteristic of the main lexeme of the domain, 重 *zhòng*.[7] Instead, it acquires the meaning 'important', as illustrated in (15).

(15) 邱　教授　　把 金钱 名利　　　　　看得　很　淡，　把
Qiū jiàoshòu   bǎ jīnqián mínglì   kàn de hěn dàn,   bǎ
Qiu professor BA money fame_and_wealth look EV very insipid BA
责任感　　　　　　却 看得　很　重。
zérèngǎn   què kàn de hěn **zhòng**
sense_of_responsibility but look EV very heavy
'Professor Qiu is indifferent to money and fame, but he considers the sense of responsibility to be very important'.   (ccl)

The morpheme 重 *zhòng* can carry the meaning 'important' also when it is a constituent part of collocations and compounds: 重视 *zhòngshì* 'attach great importance' (lit. 'heavy-to look'), 重点 *zhòngdiǎn* 'key point' (lit. 'heavy-dot').

Another figurative meaning of 重 *zhòng* is that of a large amount (usually referring to money). Consider the Example (16).

(16) 重金礼聘 *zhòng jīn lǐ pìn* 'employ sb. with a high pay' (lit. 'heavy gold polite employ')
重赏 *zhòngshǎng* 'high reward' (lit. 'heavy-reward')

In European languages, lexemes with the meaning 'heavy' can also apply to nouns denoting money, although such combinations receive negative connotations in most cases, e.g. the English *heavy taxes* or French *lourds impôts* 'burdensome (lit. heavy) taxes'. Similarly, in Russian, the collocation *tjažēlye den'gi* '(lit.) heavy money' not only designates a large amount of money but also, metonymically, the money that was earned through hard work.

Similarly to the other languages in the Ryzhova et al.'s (2019) sample, the Mandarin morpheme 重 *zhòng* 'heavy' can function as an intensifier. Typologically, such intensifiers are mostly used for negatively evaluated phenomena, e.g. English *heavy calamity*, Turkish *ağır ceza* 'heavy punishment', Russian *tjažēlaja bolezn'* 'severe disease'. 重 *zhòng* occurs in such contexts as well, but – quite unexpectedly from the typological point of view – it is also attested in neutral sentences and even in contexts with unequivocally positive connotations, such as (17) and (18).

7.　Apart from 重 *zhòng*, Chinese has another monosyllable which corresponds to 'heavy' – 沉 *chén* – used predominantly in the Beijing, the Xi'an, and the Jinan dialects (HFC. 1995). Its primary meaning is 'to drown'. It acquired the meaning 'heavy' relatively late (during the Tang period, 7–8 cent.), and the majority of its figurative meanings are connected to the primary meaning ('to drown'). The disyllabic 沉重 *chénzhòng* 'heavy' is mainly used in figurative meanings with negative connotations. Since its literal meanings are rare, the link between the figurative and the literal meanings cannot be traced.

(17) 他 对 我 的 恩情 很 **重，** 我 这 辈子 都 　　 报答
tā duì wǒ de ēnqíng hěn zhòng, wǒ zhè bèizi dōu bàodá
he DIR I ATR good very heavy I this life whole thank
不 了。
bù liǎo
NEG possible
'His kindness towards me is immense, I cannot repay him in my whole life'.

[leeds]

(18) 问 他 的 职业 　　 无异 　 于 迫 他 自 承
wèn tā de zhíyè wúyì yú pò tā zì chéng
ask he ATR profession not differ from urge he self admit
失业， 　　　 对于 　　　 自尊心 很 **重** 的 人 是
shīyè duì yú zìzūnxīn hěn zhòng de rén shì
lose employment with_regard_to self-respect very heavy ATR person be
不 大 好的
bú dà hǎo de
NEG big good ATR
'To ask him about his job is the same as to make him acknowledge that he's lost it, which isn't good for a man of high (lit. heavy) self-esteem.'

Such variance in figurative usage suggests that literal meanings within the domain HEAVY may also be heterogeneous. A closer look at the literal meanings reveals that, although typologically HEAVY is usually expressed by one dominant lexeme, its origin and the contexts in which it occurs do not exactly coincide across languages.

For example, the Russian term *tjaželyj* and the English term *heavy* that have developed negative connotations originate from verbs of motion, *tjagat'* 'to pull' and *to heave*, respectively. Obviously, moving a heavy object is a physically challenging task for humans, which evokes negative connotations.

The Mandarin term 重 *zhòng* 'heavy' is related to the nominal meaning 'weight', e.g. 举重 *jǔ zhòng* ('lift' + 'heavy') 'weight-lifting'. Such words can be found in other languages as well; for instance, consider the English *weighty* or the Russian *veskij, vesomyj, uvesistyj* (all of which mean 'weighty') that can be traced back to the idea of weight (as a neutral characteristic of an object); these terms develop figurative meanings of importance and significance, e.g. Rus. *veskij argument* 'strong argument' (lit. 'weighty argument'). Since the "objective" characteristic of an object is not fraught with negative evaluation (e.g. greater weight can signify greater value of an object), the emergence of positive connotations is perfectly explainable.

The distinction between the neutral and the negatively loaded words of the HEAVY domain can also manifest itself in interrogative constructions of the 'How much do you weigh?' type (Ryzhova et al. 2019). These constructions do not contain any evaluation, so they allow only of evaluatively neutral adjectives, such as the Mandarin 重 *zhòng*, cf. (19).

(19)   你   有   多   重?
       nǐ   *yǒu*   *duō*   ***zhòng?***
       you   have   how_much   heavy
       'How much do you weigh?'

In other languages where 'heavy' is loaded with distinct negative connotations, such as Russian or English, this question can be formulated with 'heavy' only in specific contexts which are limited in scope and usage:

(20)   ?*Naskol'ko*   *ty*   ***tjaželyj?***
        how_much   you:NOM   heavy:NOM.M.SG

(21)   ?*How heavy are you?*[8]

Thus, the figurative uses of the Mandarin lexeme from the domain of HEAVY led us to reveal an important opposition between its literal meanings, corroborating the claim of MLexT that meaning extensions originate from frames.

4.   Applying typological data to lexicology

So far in this chapter we have presented several semantic domains and the associated phenomena that are peculiar to Mandarin.

In particular, potential oppositions that otherwise could only be hypothesised from indirect evidence are realized in the Chinese language at the lexical level. For example, in the domain ROUGH, one might predict a lexical opposition for expressing the roughness of surfaces of artefacts versus surfaces of natural objects, since such an opposition for the object type is attested in the antonymous domain SMOOTH (see Figure 2 above). This hypothesis was empirically confirmed only when this opposition was found in Mandarin Chinese, where the lexeme 粗糙 *cūcāo* can describe both classes of objects, while 毛糙 *máocao* can only apply to artefacts (see Chapter 6, this volume).

At the same time, typological data in its turn offers valuable insights which can contribute to understanding of the system of lexical oppositions in Mandarin Chinese. The most promising practical application of this alliance appears to lie within the field of lexicography.

The methods of semantic analysis developed within the framework of MLexT, as well as in lexical typology in general, are capable of effectively distinguishing the

---

8.   *How much do you weigh?* would be preferable here. Speakers of American English report that the question *How heavy are you?* can presuppose something like *I see that you're quite heavy. Just how heavy are you?* (http://forum.wordreference.com).

meanings of seemingly synonymous lexemes. This is due to the fact that the sets of situations in which synonyms (even close ones) can occur are usually not identical. The opportunity to inspect the patterns of lexical distribution in one language and compare them with other languages promotes a clearer and more systematic description, and reduces the subjectivity of analysis which is almost inevitable when a researcher works with only one language.

The typological approach also makes it possible to depart from the commonly practiced yet somewhat problematic technique of defining word meanings through synonyms, which often results in circular definitions. Let us illustrate this problem by definitions of several synonymous lexemes that share the meaning of 'smooth' (as presented in the *Dictionary of Modern Chinese* (XHC 6. 2015)):

A. 滑 *huá* 'slippery, smooth': 光滑 *guānghuá* 'smooth, sleek', 滑溜 *huáliu* 'slippery, smooth'
B. 光溜 *guāngliu* 'smooth, glossy': 光滑 *guānghuá* 'smooth, sleek', 滑溜 *huáliu* 'slippery, smooth'
C. 光滑 *guānghuá* 'smooth, sleek': object surface is 平滑 *pínghuá* 'smooth, flat', not rough
D. 滑溜 *huáliu* 'slippery, smooth': 光滑 *guānghuá* 'smooth, sleek' (with a positive connotation)
E. 平滑 *pínghuá* 'smooth, flat': 平 *píng* 'level' and 光滑 *guānghuá* 'smooth, sleek'

It can be seen that, according to these definitions,

(A) = (B) = > (C) = > (E) = > (C)
$\qquad$ = > (D) = > (C),
where "=" denotes the equality of definitions, and
"A = > B" means that the definition of lexeme A contains lexeme B.

Obviously, such definitions only obscure the differences among the meanings of the lexemes, especially to a user who is not a native speaker of Mandarin Chinese. The popular dictionaries of synonyms (Zhang & Zhang 2005; Cheng 2010; Zhu 2009; Liu 1987) that could have shed more light on this issue, do not contain the lexemes in question. We believe that describing the co-occurrence patterns and the situations in which lexemes are used (i.e. frames) may be a more effective approach to designing dictionary entries.

In addition, the data obtained in lexico-typological studies of semantic domains can be used to compile or augment dictionaries of synonyms. For example, when collecting the material for this study, we noticed that the existing dictionaries of synonyms of Mandarin Chinese tend to group together only the tokens that match in number of syllables.

Consider the dictionary of synonyms edited by Cheng Rong (2010), which is the most extensive of all the currently available dictionaries of this type; it lists disyllabic lexemes with their disyllabic synonym counterparts only. For example, in the domain SHARP, the dictionary totally neglects the monosyllabic 尖 *jiān* 'sharp, pointed', while scrutinising a large group of disyllabic lexemes such as 尖利 *jiānlì* 'sharp, shrill, piercing', 锋利 *fēnglì* 'sharp, incisive', 尖锐 *jiānruì* 'sharp-pointed, shrill, acute', 锐利 *ruìlì* 'sharp, penetrating', 犀利 *xīlì* 'sharp; incisive; trenchant'. Similarly, when presenting the semantic domain SOFT, the monosyllable 软 *ruǎn* 'soft' is not mentioned, while disyllables are presented in abundance, cf.: 软和 *ruǎn-huo* 'soft, gentle, kind', 绵软 *miánruǎn* 'soft, weak', 柔软 *róuruǎn* 'soft, flexible', 酥软 *sūruǎn* 'soft, limp, weak', 柔软 *róuruǎn* 'soft, flexible'.

As a result of this artificial bias, monosyllabic lexemes tend to be largely overlooked by authors of synonym dictionaries. Yet, monosyllabic lexemes are rightful members of the lexicon, and their difference from disyllables goes beyond mere rhythmic constraints. For example, in the domain SOFT, the disyllabic token 柔软 *róuruǎn* 'soft' cannot be used interchangeably with 软 *ruǎn* in contexts like 软石头 *ruǎn shítou* 'soft stone (material)' or 软水果 *ruǎn shuǐguǒ* 'soft fruit', since its meaning has an additional shade of flexibility, resilience, and the ability to restore the original shape. Speaking of the domain SHARP, the meaning conveyed by the monosyllable 尖 *jiān* 'sharp, pointed' to describe sounds also differs from the disyllable 尖利 *jiānlì* 'sharp, shrill, piercing': the former describes a high-pitched thin voice (neutral characteristic), while the latter describes a shrill and unpleasant voice (negative characteristic).

Thus, systematic application of the methods of lexical typology can be instrumental in improving the quality of monolingual Chinese dictionaries.

5.   Conclusion

In this chapter we have shown that, on the one hand, the parameters of typological variability and the models of semantic shifts developed within the MLexT framework are applicable to Mandarin Chinese. On the other hand, the Chinese lexicon demonstrates a number of features that appear unexpected from a Eurocentric point of view. Yet, a closer look at the European languages reveals that such seemingly exotic effects are in fact equally characteristic of them as well, albeit they may be more peripheral and could have remained unnoticed had it not been for the comparison with the Mandarin data.

Furthermore, we claimed that lexico-typological methodologies can be effectively introduced into the practice of Chinese lexicography. Lexical typological

methods and their results can prove instrumental in detecting and improving problematic parts of lexicographic descriptions, since they equip the lexicographer with tools to differentiate lexemes that otherwise would seem to be fully synonymous.

Abbreviations

ACC	accusative		IMP	imperative
ATR	attributivizer		INS	instrumental
BA	preposition, putting the object to the position before the verb		LOC	locative
			N	neutral
CLF	classifier		NEG	negation
EV	structural particle, used after a verb, linking it to following phrase indicating effect, degree, possibility, evaluation etc.		NOM	nominative
			PFV	perfective
			PL	plural
F	feminine		PST	past
GEN	genitive		SG	singular

Literature

Apresjan, Ju. 2000. *Systematic Lexicography*. Oxford: Oxford University Press.

Cheng, Rong. 2010. 程 荣，同义词大词典(辞海版)，上海辞书出版社. [Big Synonym Dictionary (Cihai edition)]. Shanghai: Shanghai Dictionary Publishing House.

Feng, Shengli. 2009. 冯胜利 汉语的韵律、词法与句法 [Interactions between Morphology, Syntax and Prosody in Chinese]. Beijing: Peking University Publishing House.

Feng, Shengli. 2015. 冯胜利. 语体语法的逻辑体系及语体特征的鉴定 [Logical Relations in the Stylistic Grammar and the Differentiation of Stylistic Features]. 汉语应用语言学研究. [Studies in the Applied Linguistics of Chinese]. Beijing: Commercial Press.

Feng, Shengli. 2017. Disyllabification. In *Encyclopedia of Chinese Language and Linguistics. Vol. II*, 108–113. Leiden, Boston: Brill.

François, A. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic* [Studies in Language Companion Series 106], M. Vanhove (ed), 163–215. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.106.09fra

Georgakopoulos, T. & Polis, S. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2): 1–33. https://doi.org/10.1111/lnc3.12270

Haspelmath, M. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In *The new psychology of language, vol. 2*, M. Tomasello (ed), 211–242. Mahwah, NJ: Lawrence Erlbaum.

HFC 1995.《汉语方言词汇》北京大学中国语言文学系语言学教研室主编, 语文出版社. [Lexicon of Chinese Dialects by Chinese Department of Peking University]. Beijing: Yuwen Publishing House.

Jiang, Shaoyu. 2015. 蒋绍愚 汉语历史词汇学概要. 北京：商务印书馆 [The Foundations for the Diachronic Study of the Chinese Lexicon]. Beijing: Commercial Press.

Ke, Hang. 2012. 柯航 现代汉语单双音节搭配研究. 北京：商务印书馆. [A Study of the Co-occurrence Potential of Mono- and Disyllabic Lexemes of Modern Chinese]. Beijing: Commercial Press.

Kyuseva, M.V., Ryzhova, D.A. & Kholkina, L. S. 2012. Prilagatel′nye «legkij» i «tjaželyj» v tipologičeskoj perspektive. [The adjectives 'light' and 'heavy' in a typological perspective]. In *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" -2012*, 247–255. Moscow: RGGU.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.
https://doi.org/10.7208/chicago/9780226471013.001.0001

Lakoff, G. & Johnson, M. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.

Liu, Shuxin. 1987. 刘叔新, 现代汉语同义词词典, 天津人民出版社. [Synonym Dictionary of Contemporary Chinese]. Tianjin: Tianjin Peoples' Publishing House.

Pan, Wenguo, Ye Buqing & Han Yang. 1993. 潘文国，叶步青，韩洋. 汉语的构词法研究, 台湾学生书局. [Studies in the Chinese Word-formation]. Taiwan: Taiwan xuesheng shuju.

Pavlova, E.K. 2014. Leksiko-tipologičeskoe opisanie priznakov 'mjagkij'-'tverdyj' v jazykax raznyx semej [Lexical-typological description of the qualitative features 'soft' – 'hard' in the languages of various phyla]. Diploma paper, Moscow State University.

Pavlova, E. K. & Kholkina, L. S. 2013. Words of hardness and softness: towards lexical typology. In Association for Linguistic Typology. The 10th Biennial Conference (Leipzig, 2013, August 15–18). Abstracts, <http://www.eva.mpg.de/lingua/conference/2013_ALT10/pdf/abstracts/Abstracts_ALT10_complete.pdf>.

Rakhilina, E. V. & Reznikova, T. I. 2016. A Frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–129. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110377675-004

Rakhilina, E. V., Ryzhova, D. A. & Badryzlova, Yu. G. In press. Lexical typology and semantic maps: perspectives and challenges. *Zeitschrift für Sprachwissenschaft*.

Ryzhova, D., Rakhilina, E. & Kholkina, L. 2019. Approaching perceptual qualities: the case of heavy. In *Perception Metaphors*, L. J. Speed, C. O'Meara, L. San Roque & A. Majid (eds), 185–208. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.19.10ryz

Semenas, A.L. 1992. *Leksikologija sovremennogo kitajskogo jazyka* [Modern Chinese Lexicology]. Moscow: Nauka.

XHC 6. 2015. 现代汉语词典. 北京: 商务印书馆 [Dictionary of Contemporary Chinese]. Beijing: Commercial Press.

Ye, Wenxi. 2004. 叶文曦. 汉语语义范畴的层级结构和构词的语义问题. 语言学论丛(第二十九辑) [The layer structure of the semantic categories of Chinese and issues of word-formation semantics]. *Yuyanxue luncong* 29: 95–109.

Zalizniak, Anna A., Bulakh, M., Ganenkov, D., Gruntov, I., Maisak, T. & Russo, M. 2012. The Catalogue of Semantic Shifts as a Database for Semantic Typology. *Linguistics* 50(3): 633–669.
https://doi.org/10.1515/ling-2012-0020

Zhang, Zhiyi & Zhang Qingyun. 2005. 张志毅, 张庆云, 新华同义词词典, 商务印书馆. [Xinhua Synonym Dictionary]. Beijing: Commercial Press.

Zhu, Jingsong. 2009. 朱景松, 现代汉语同义词词典, 语文出版社 [Synonym Dictionary of Contemporary Chinese]. Beijing: Yuwen Publishing House.

CHAPTER 10

The qualitative lexicon in Russian Sign Language from a typological perspective

Maria Kyuseva
University of Surrey

The paper describes several expressions of physical qualities (namely, SHARP, BLUNT, OLD, THICK-WIDE, and THIN-NARROW) in Russian Sign Language (RSL) from a lexical typological perspective. This is the first study to analyse a sign language from the standpoint of the MLexT framework. The results show that RSL structures the aforementioned domains on the same grounds as spoken languages. This finding strongly supports the hypothesis that lexical systems are cognitively motivated. At the same time, RSL exhibits a number of non-trivial lexicalization strategies, which are not accidental but can be explained by the fact that this language is produced in the visual modality.

Keywords: Russian Sign Language, RSL, sign language lexicon, lexical typology, physical qualities

1. Introduction

This paper presents an analysis of signs denoting physical qualities in Russian Sign Language (RSL), namely, SHARP, BLUNT, OLD, THICK-WIDE, and THIN-NARROW, and compares them with spoken languages. A great number of typological studies that include data of both signed and spoken languages in their samples are generally limited to grammatical phenomena (e.g., Pfau & Steinbach 2006; Sandler & Lillo-Martin 2006; Zeshan 2006). The relationship between auditory-oral versus visual-gestural modality in communicative systems at the lexicon level still remains largely understudied (compare, though, Grose 2012; Sagara & Zeshan 2016) and has never been investigated within the MLexT framework.

Such a comparison invokes a number of potential problems, the main one being a lack of the full correspondence between spoken words and signs of sign languages (SLs). Although the methodology of lexico-semantic analysis, described in this volume, has already been shown to be suitable for comparison of spoken

languages with very diverse structures, it has not yet been validated against sign language data. The secondary research question of this study, then, is: to what extent is the MLexT methodology applicable to sign languages?

The rest of the paper is organized as follows: Section 2 introduces the problem of the word in sign languages and discusses the nature of iconicity; Section 3 reviews previous comparative works on the lexicons of signed versus spoken languages; finally, in Section 4 we present an analysis of RSL physical qualities in a typological perspective.

2. Iconicity and the word in sign languages

The main problem of signed versus spoken language comparison derives from iconicity, an inherently more pervasive property of the former. Sign languages have been repeatedly stated to exhibit a much higher degree of iconicity than spoken ones (Meir 2010; Sandler & Lillo-Martin 2001; Taub 2004). It is connected with the fact that the major portion of information in both hearing and deaf communities is transmitted and perceived through the visual, not the auditory, channel (Taub 2004). Thus, there are many more objects suitable for iconic encoding in sign than in speech. Sign languages, produced in the visual modality, possess a developed apparatus for iconic encoding of objects. Pietrandrea (2002), for example, has found that 50 percent of Italian Sign Language hand configurations and 67 percent of the occurrences of body locations in signs are iconic. Spoken language iconicity, on the contrary, is limited largely to onomatopoeia and some other disparate phenomena; see, however, Blasi et al. (2016); Perlman et al. (2018), and Perniss et al. (2010) who argue for a larger degree of iconicity in spoken languages than is usually assumed.

A simple example of sign language iconicity at the lexicon level is the RSL sign WALL[1] (Figure 1). The hand configuration in this sign depicts the prototypical shape of the wall – flat and thin; the place of the hand depicts its default location near the signer; finally, the direction of the movement and orientation iconically depict the vertical orientation of the wall.

---

1. In this article, sign language signs are marked with small caps and italics; semantic domains are marked with small caps; frames and meanings of words/signs are marked with single quotes, as in the example: The RSL sign GREEN from the semantic domain COLOR has a metaphorical meaning 'young'.

Figure 1. RSL sign WALL [RSL corpus][2]

The extensive role of iconicity was partly responsible for the lack of recognition of the status of sign languages as fully functional independent natural communicative systems. It took several decades of studies to provide compelling evidence for this point of view (see the classical works Stokoe 1960; Bellugi & Klima 1979, but also Sandler & Lillo-Martin 2001). Now it is widely accepted that iconicity does not prevent sign languages from being "real" languages, but that it is present in the whole system and is connected with fundamental differences between signed and spoken languages. It is primarily iconicity that is responsible for the heterogeneity of sign language lexicons, for which it is hard to find any parallel in spoken languages.

A rough equivalent of the spoken language word in sign languages is the sign. It consists of smaller structural elements. According to the traditional classification stemming from Stokoe's works, these are: handshape, orientation, place of articulation, movement, and non-manual features. The non-manual component includes two groups of mouth actions: mouthing and mouth gestures (Boyes-Braem & Sutton-Spence 2001). Mouthing is the silent articulation of the correspondent spoken language word, whereas mouth gestures are generally believed to be sign-language-internal elements. Examples of the latter are: /af/, /mmm/ (pressed lips), /lm/, /th/. Table 1 presents values of these parameters for the RSL sign WALL (Figure 1).

Table 1. Sign components of the RSL sign WALL

handshape	
orientation	horizontal, palm towards the signer
place of articulation	neutral space in front of the signer
movement	upward, straight
non-manual component	*stena* (lit. 'wall'; mouthing)

2. http://rsl.nstu.ru/data/view/id/249/t/28834/d/29754j

As sign components can form minimal distinctive oppositions, they are frequently compared to spoken language phonemes.[3] Consider RSL pairs MAN vs. WOMAN and TUESDAY vs. WEDNESDAY (Figures 2 and 3). The former two signs differ in place of articulation, the latter ones differ in hand configuration.

a.   b.
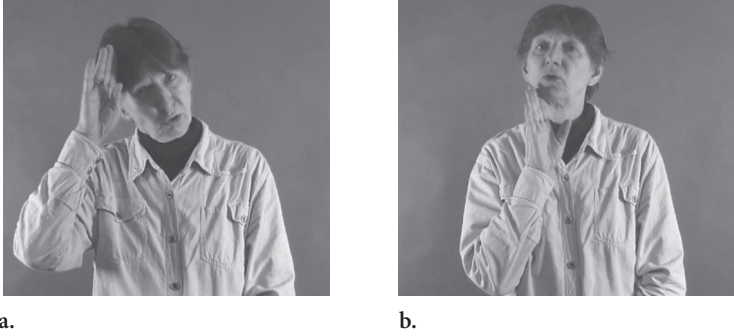
Figure 2. RSL signs MAN (a) and WOMAN (b)

a.   b.

Figure 3. RSL signs TUESDAY (a) and WEDNESDAY (b)

Signs which have conventional meanings in a signing community and whose components remain unchanged in different contexts belong to the "core" domain of the lexicon (Johnston & Schembri 2007). An example of a core RSL sign is TUESDAY (Figure 3a): its form is the same in different contexts, such as: 'I'll come on Tuesday', 'Tuesday is the second day of the week', 'I hate Tuesdays'. If a sign has a given form and meaning only in a given context, it belongs to the "non-core" (Johnston & Schembri 2007), or "productive" (Brennan 1990) lexicon. Consider, for example, the meaning 'thick' in RSL, which is expressed with different forms in the phrases 'thick wall', 'thick tree', 'thick rope', shown in Figure 4.

---

3.   An important difference between spoken language phonemes and sign components is that the former generally have the form of fairly well-distinguished segments, whereas the latter are expressed simultaneously.

a.   b.   c.
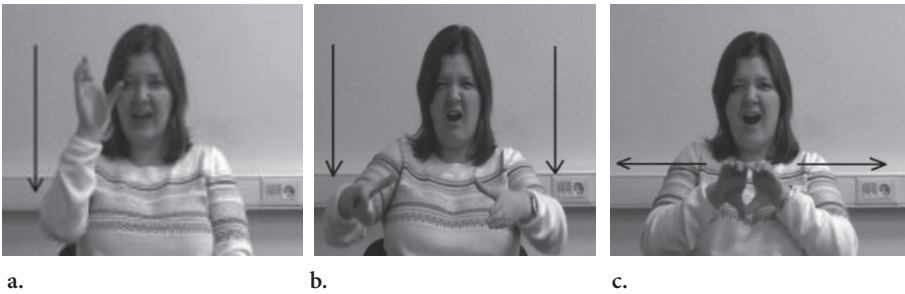
Figure 4. RSL 'thick' in the context of the nouns 'wall' (a), 'tree' (b), 'rope' (c)

In Figure 4a, one hand with its palm oriented away from the signer moves vertically down; in Figure 4b, not one, but two hands move down with the palms oriented towards the midsaggital plane; in Figure 4c, the hands move horizontally away from each other. Such context-dependent changes can happen only with iconic sign elements. In the discussed examples, the sign movement iconically outlines the form of the described object. Since there is a potentially infinite number of object shapes, the list of possible sign forms describing these shapes is also non-exhaustive.[4] That is why non-core signs are sometimes compared to pantomime and gesticulation of non-signers (Kendon 2004). These signs violate some well-formedness constraints (Johnston & Schembri 2007), they are believed to be actively created by signers from combinations of meaningful units rather than kept in the mental lexicon (Brennan 1990), and their status as "words" remains unclear (Grose 2012).

3.   Sign vs. spoken language lexicon: Previous studies

The presence of pantomime-like forms in sign languages poses significant challenges for the comparative analysis of the lexicon. Several successful attempts include lexical typological studies of color terms, kinship terms and numerals (see, among others, Grose 2012; Sagara & Zeshan 2016; de Vos & Pfau 2015).

Color terms represent the most extensively analyzed semantic domain in both signed and spoken languages. Sagara & Zeshan (2016) report the data from more than 33 sign languages. Their data collection techniques included a typological questionnaire and elicitation games. The typological questionnaire showed little resemblance to those used in the MLexT framework and served more as a check list for researchers; see an illustration of this in Figure 5.

---

4.   See, however, Willbur (2013) who challenges the notion of non-exhaustiveness in non-core sign formation.

Which semantic types do the color terms in your sign language belong to?
The sign is semantically related to an object, e.g. a body part TEETH to indicate 'white', or ORANGE to indicate 'the color orange'. – List the signs in this category.
The sign is linked to the spoken/written language either through fingerspelling or mouthing or both, e.g. ASL YELLOW. – List the signs in this category.
The sign is non-iconic. – List the signs in this category.
Other, namely………………………………………… – List the signs in this category.

Figure 5. The typological questionnaire used in Sagara & Zeshan (2016) for color term collection[5]

In the Sagara & Zeshan (2016) study, elicitation games represented a direct data collection method and included a director-matcher color game, played by two participants. The director described to the matcher the colors on a colored picture, and the matcher's task was to arrange the correct color chips on an equivalent black-white picture. These techniques resulted in color term lists for 33 sign languages. After eliminating borrowings, metonymic terms (such as the sign *TEETH* to indicate white) and morphologically complex signs, the experts formed lists of *basic* color terms, which were then analyzed against the color term hierarchy introduced in Berlin & Kay (1969) for spoken languages (see Figure 6).

black/white → red → yellow/green → blue → brown → other

Figure 6. Berlin & Kay (1969) color term hierarchy[6]

The headed arrows indicate a hierarchical relationship among terms. The hierarchy predicts that if a language has a basic term for a color placed on a given point of

---

5.   http://www.uclan.ac.uk/research/explore/projects/sign_language_typology.php

6.   The slash between the colors 'black' and 'white' in the hierarchy operates as the logical conjunction: according to Berlin and Kay (1969), all languages have basic terms for both 'black' and 'white'.

The slash between the colors 'yellow' and 'green' in the hierarchy operates as the logical disjunction: if a language has a term for one of these colors, it might have or not have a term for the other.

the continuum, it will also have basic terms for all the colors placed to the left of this point on the hierarchy. For example, if a language has a basic term for 'green', it will also have 'red', 'black' and 'white' basic terms.

As the Sagara & Zeshan (2016) study showed, by and large, the Berlin & Kay hierarchy applies to sign languages. For example, Adamorobe Sign Language has single manual signs for three colors: 'black', 'white', and 'red'; Kata Kolok has four lexical signs designating 'white', 'black', 'red', and 'blue-green'; Yolngu Sign Language, in use among an Aboriginal community in Australia, has six color terms, none of which are from the far right end of the hierarchy. Note, however, a famous counter-example from Inuit Sign Language, which has only two lexical signs, BLACK and RED, and does not have a sign for 'white'; de Vos & Pfau (2015).

An interesting peculiarity of sign languages is *initialization*, which is a strategy of borrowing unique to the visual modality. It consists in the production of a sign with a handshape corresponding to the first letter of the color word in the surrounding spoken language (e.g., 'G' for 'green' and 'P' for 'purple' in American Sign Language). In line with Berlin & Kay's hierarchy, initialized signs typically occur in the right end of the continuum.

The study by Sagara and Zeshan (2016) confirms the principal possibility of including sign language data in a lexical typological analysis. The existing comparative analyses of kinship terms and numerals in sign versus spoken languages come to the same conclusion; see Grose (2012); Wilkinson (2009); Woodward (1978); Zeshan et al. (2013). However, all these studies are limited to the "core" signs, which are conventionalized in Deaf communities, fully specified and interpreted unambiguously without context. Therefore, the question of whether the signed versus spoken language comparison is possible against the whole range of lexical items (including non-core signs) remains unanswered.

4. Physical qualities in Russian Sign Language

The present study is an analysis of signs denoting physical qualities in Russian Sign Language in comparison with spoken languages. The qualities we chose for the study are semantically diverse; they can be divided into three groups:

– functional (SHARP, BLUNT)
– temporal (OLD)
– visual (THICK-WIDE, THIN-NARROW)

We expected that semantically different domains would be covered by structurally different signs and, consequently, that we would be able to test our methods on not only core, but also peripheral sign language words.

4.1 Research methodology

We used a technique traditional for the MLexT group research: studying collocations by filling out lexico-typological questionnaires (see Chapter 1 in the present volume). As the only existing corpus of Russian Sign Language (www.rsl.nstu.ru) is not large enough to enable a study of the lexicon, we relied primarily on elicitation sessions for filling out the questionnaires and used corpus data as supplementary material only.

We had two kinds of questionnaires: a list of Russian phrases of the form 'quality word+ name of the object' and a list of sentences containing phrases in question. See, for instance, a fragment of different questionnaire forms for the semantic domain SHARP in Figure 7.

ostryj nosh 'sharp knife'	*Mne nyzhno kupit' ostryj nozh dlya pokhoda* 'I need to buy a sharp knife for hiking'
ostroe kopje 'sharp spear'	*Nigde ne videl takogo ostrogo kopja* 'I have never seen such a sharp spear'
ostryj lokot' 'sharp elbow'	*Menja zhenschina tolknula v avtobuse ostrym loktjem* 'A woman pushed me with her sharp elbow on the bus'
ostryj klyk 'sharp fang'	*Bud' ostorozhen! Y etoj sobaki ostrye klyki* 'Beware, this dog has sharp fangs'
ostryj kolpak 'pointed cap'	*Mne mama kuoila ostryj kolpak dlya spektaklya* 'My mother bought me a pointed cap for the performance'
ostraya kosa 'sharp scythe'	*Khorosho kosit' travu ostroj kosoj* 'It's nice to mow grass with a sharp scythe'

Figure 7. SHARP: Fragment of the questionnaire

Native signers translated these Russian stimuli into Russian Sign Language. Each questionnaire was filled out by five to seven informants. We analyzed the video recordings of the sessions using the ELAN 4.9.0 software (http://tla.mpi.nl/tools/tla-tools/elan/). The annotation included sentence translations, as well as the right and the left hand glosses for each sign. Additionally, a phonetic form of the signs denoting physical qualities in question was described with the help of the HamNoSys notation (Hanke 2004; Prillwitz et al. 1987). Further analysis consisted in outlining the set of signs covering the semantic domains and their semantic mapping.

4.2 Results of the study

The analyzed qualitative domains in RSL fall into two groups: one with core signs only and one which included (but was not limited to) non-core signs, or, more precisely, the class of non-core signs called "size and shape specifiers", or SASSes, in the literature (Supalla 1986; Zwitserlood 2003). The first group allowed for a

straightforward application of our lexical typological framework; the second one required some assumptions about the organization of the sign language lexicon before application of the analysis. Let us start with an easier case and then move to a more problematic one.

4.2.1   *Semantic domains with core signs only*

Two out of seven qualities investigated in Russian Sign Language are covered by core signs exclusively: BLUNT and OLD. Blunt objects are described in RSL with one sign (*BLUNT*; Figure 8), and old people/objects/events are described with two simple signs (*OLD* as in Figure 9a; *PREVIOUS* as in Figure 9b) and one compound (*OLD+LONG.AGO*; as in Figure 9c).

Figure 8. RSL sign *BLUNT*

a.   b.

c.
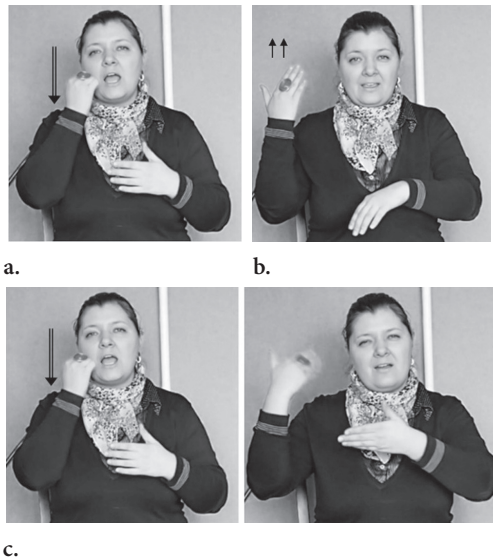
Figure 9. RSL signs *OLD* (a), *PREVIOUS* (b), *OLD+LONG.AGO* (c)[7]

---

7.   Double-lined arrows in (a) and (c) stand for downward movement: solid arrows in (b) denote movement towards the signer.

These systems can be easily compared with the lexical items from corresponding semantic domains in spoken languages and placed onto semantic maps. Thus, the domain BLUNT, reported on in the analysis of twenty one spoken languages by Kyuseva et al. (Chapter 2 of this volume), has two semantic frames[8] in the zone of physical meanings: 'instruments that cannot cut well' (*blunt knife/saw/scythe*) and 'instruments that cannot pierce well' (*blunt needle/spear/awl*). Three basic lexicalization strategies are registered in spoken languages for expressing these two frames: (i) each frame is covered by its own adjective (e.g. Finnish; Komi, Uralic), (ii) none of the frames is covered by a 'blunt' lexeme (e.g. French, where these meanings are expressed with either a periphrasis or a participle); and (iii) both frames are described with one word. The last strategy is the most frequent in the study, and is attested in Russian, Serbian, English, Chinese, Basque, German, Japanese, Korean, and Malay. It also applies to the RSL system, as BLUNT describes both types of instruments (Figure 10).
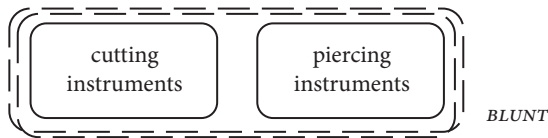
Figure 10. Semantic domain BLUNT in RSL

As for the domain OLD, RSL exhibits a distributive strategy of lexicalization: OLD describes elderly people and objects with an expiring usage period ('old person', 'old cloth'); PREVIOUS denotes replaced objects and individuals ('former director'); and the compound OLD+LONG.AGO covers the frame of people and objects from the past era ('antique coin'). This is displayed in Figure 11.
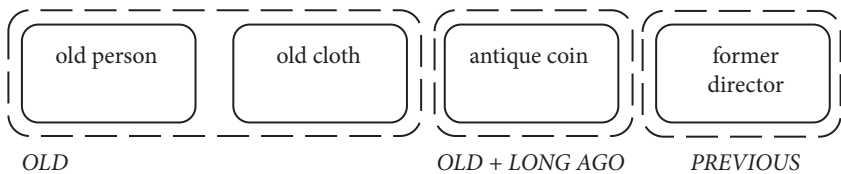
Figure 11. Semantic domain OLD in RSL

---

8.   I use the term "frame" in Rakhilina & Resznikova's (2016) sense as a prototypical situation that can be described in a language with a separate lexeme. This term used in the same way throughout this volume.

In contrast with the domain BLUNT, the RSL lexicalization for OLD is extremely rare in spoken languages (see Chapter 10 in this volume). In spoken languages, it is much more frequent to find one of three systems: (i) dominant (one adjective describes all types of objects); (ii) binary (there are two lexemes in the domain); or (iii) distributive (each frame is covered by its own lexeme; or there are three lexemes in the domain, where the first lexeme covers the frame 'old person', the second lexeme covers the frame 'old cloth', and the third lexeme covers the frames 'antique coin' and 'former director'). Therefore, although RSL does not violate any generalisation in this semantic zone, it shows a certain peculiarity.

4.2.2 Semantic domains with core and peripheral signs

The remaining qualities we examine in this paper are covered by both core and peripheral signs. These semantic domains are: SHARP, THICK-WIDE, and THIN-NARROW. Consider the domain SHARP. Its semantic organisation is illustrated in Figure 12 (based on Chapter 2 of this volume). Two out of three frames describe well-functioning instruments: cutting (e.g., *sharp knife, saw, scythe*) and piercing (e.g., *sharp needle, spear, arrow*) ones. The last frame is visual and it denotes objects with a pointed end (*pointed nose, elbow, horn*, among others).

cutting instruments	piercing instruments	pointed objects
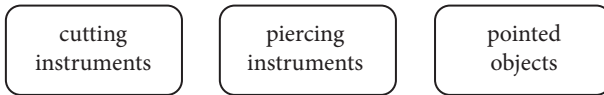
Figure 12. Semantic map of the domain SHARP

In RSL, the frame 'cutting instruments' is covered by the core sign *SHARP* (Figure 13). This sign, for example, is used for describing a sharp blade or a sharp sword; it will express the quality in question in the sentences 'I cut myself with a sharp knife'; and 'You should better use a sharp saw here'.

Figure 13. RSL sign *SHARP*

The frames 'piercing instruments' and 'pointed objects' are described in RSL with highly iconic non-core sign forms. As Figure 14 illustrates, different referents require different forms to express the meaning 'sharp/pointed'. In 14a, the active hand is moving away from the elbow and is changing its configuration from 🤚 to ☝; in 14b, the active hand in the configuration 👆 moves away from the passive hand's index finger while closing its own index and thumb; finally, in 14c, the hand is located on the signer's head, it moves away from it in the arcing trajectory while bringing the fingers and the thumb together. As we can imagine an infinite number of pointed objects with different positions in space and different aspects of shape, so the number of sign forms in this zone is infinite as well.

a.

b.

c.
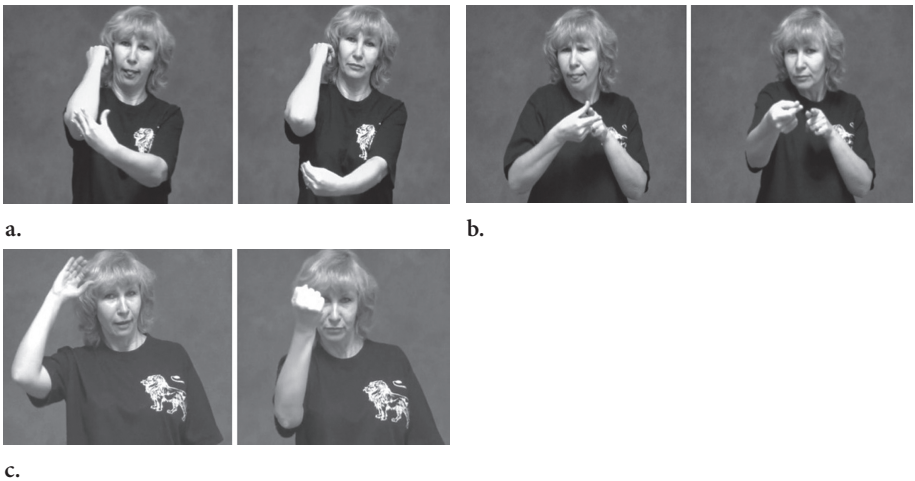
Figure 14. Peripheral signs of the semantic domain SHARP in different contexts: 'elbow' (a), 'knife' (b), 'horn' (c)

Non-core signs do not have an established "word status" in sign language linguistics. Accounts of these signs range from fully-morphemic to non-linguistic ones. According to the fully-morphemic view (Supalla 1986; Benedicto & Brentari 2004), these signs are composed out of a closed list of morphemes; each morpheme describes one aspect of the object's shape (i.e., dimensionality, roundness, width, length, orientation, etc.). The combination of the meanings of these morphemes results in a complex depiction of the object. Let us call this view the "linguistic" interpretation of non-core sings; according to it, these signs constitute part of the linguistic system. An opposite interpretation states that non-core signs are part of a different semiotic system. For example, Cogill-Koez (2000a, 2000b) claims that when signers use these signs, they find themselves in the schematic visual representation mode. Non-core signs, according to this view, are similar to drawings. When signers use them to depict an object's shape, they do not name the shape, but "draw" it instead. This is the "non-linguistic" interpretation of these signs. The third

widespread opinion places these signs between the "linguistic" and "non-linguistic" extremes. Liddell (1980, 2003) proposes an account, according to which these signs have two parts, i.e. the symbolic linguistic and the gestural part. The linguistic part consists of clearly defined morphemes that are kept in signers' mental lexicons and describe the world *categorically*. The gestural part reflects the characteristics of a referent in an *analogue* manner. It does not have a closed list of forms and it is not remembered by signers, but is *created* every time the sign is used. This is the "partly-linguistic" account of these signs (see (Emmorey & Herzig 2003) for an experimental testing of this hypothesis).

These interpretations of non-core signs lead to quite different analyses of lexical zones SHARP, THICK-WIDE, and THIN-NARROW. Thus, if these signs represent a different semiotic mode (let us call it "depicting" mode as opposed to "linguistic" one), then in the zones in question RSL and spoken languages are not comparable. When signers depict visual qualities, they engage fundamentally different cognitive processes than speakers. For the domain SHARP, this is illustrated in Figure 15a. The leftmost frame 'cutting objects' in RSL is covered by the core sign *SHARP* (indicated with the dashed outline). The two remaining frames do not have a linguistic expression in RSL and, therefore, are left unmarked. If non-core signs are, as the fully-morphemic account states, composed out of a closed list of morphemes, then the zones SHARP, THICK-WIDE and THIN-NARROW have much richer systems in RSL, then in analysed so far spoken languages. The lexicalization of these zones in RSL is structured according to such qualities of an object as dimensionality, roundness, orientation, width, length, etc. Accounting for this would require modification of the existing semantic maps and increasing the number of frames. See an illustration for the domain SHARP in Figure 15b.
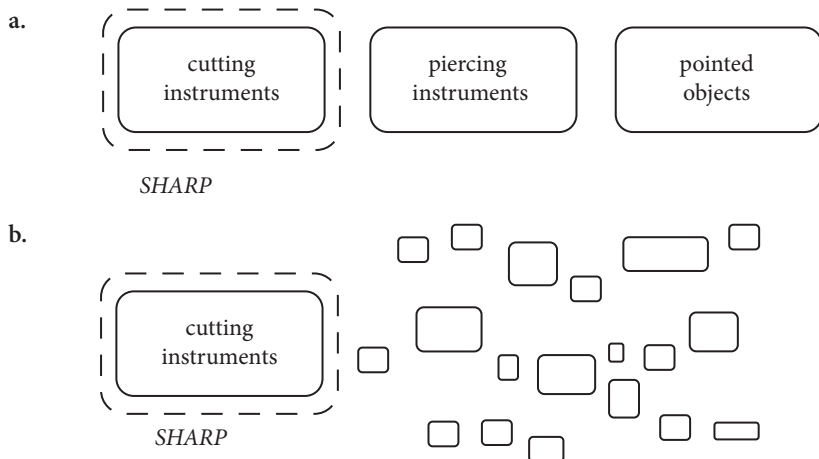
Figure 15. Possible analyses of the semantic domain SHARP in RSL: non-linguistic account of non-core signs (a) and fully-linguistic account of non-core signs (b)

The partly-linguistic account of non-core signs means that while some parts of these signs describe objects in an analogue manner, other parts are lexical and describe the quality in question categorically. These lexical sign components show the lexical organization of semantic zones. If this interpretation is correct, then the amount of frames distinguished by RSL is smaller than in Figure 15b, and the frames themselves are potentially similar to the ones distinguished by spoken languages. This interpretation of non-core signs proved to be the most suitable for accounting for the RSL data. Indeed, a thorough examination of the answers to the lexico-typological questionnaires shows that different components of non-core signs depicting pointed, thick, wide, thin, and narrow objects have unequal status. Let us discuss pointed objects first. While location, hand orientation and movement characteristics in RSL signs describing pointed objects are highly context-dependent and have different values when used with different objects, the variety of handshapes comes down to two sequences: [handshape] -> [handshape] and [handshape] -> [handshape]. The sign "->" denotes the change of configuration. Thus, the first sequence means that the hand has the configuration [handshape] at the initial stage of the sign production, and the configuration [handshape] at the final stage of the sign production. This sequence marks relatively large three-dimensional referents, such as mountains, witch hats, stakes. The second sequence means that the hand changes its shape from [handshape] to [handshape] during the sign production. It is used when the sign describes flat or very thin objects, i.e. a scythe, a shovel, or a needle.

Therefore, the lexical part of the non-core signs in the semantic domain SHARP is the handshape component; it points to one of the two lexical values and signifies the object's dimensionality. The meaning 'pointed' is expressed in the handshape change: the object in question is thicker at the bottom (open handshape) and thinner at the end (closed handshape). The gestural part contains all the remaining manual elements of the sign: orientation, location, and movement. These components express such characteristics of the object as position in space, orientation, length, and the presence/absence of bending. This analysis requires a modification of the semantic map, as in Figure 16 (modified relative to Figure 12 above).

While the left part of the map remains unchanged, each of the two rightmost frames is divided in two: three-dimensional (3D) versus flat objects. The upper row (3D) is covered by the non-core lexeme SASS(5;5#):pointed,[9] where "SASS" stands for "size and shape specifier", and (5;5#) is a letter/number representation of the start and end handshapes ([handshape] and [handshape], respectively; according to the notation

---

9.   In this notation of non-core signs, I follow the convention used in Burkova&Kimmelman (2019) handbook on Russian Sign Language research. They propose to gloss non-core signs in three parts: the first part denotes the type of the non-core sign (i.e. SASS), the second part marks the handshape(s) used in the sign production (i.e. 5; 5#); finally, the last part provides the semantic meaning of the sign.

SASS(5;5#):pointed

piercing instruments – 3D (*stake*)	pointed objects – 3D (*mountain*)

SHARP

cutting instruments (*knife*)	piercing instruments – flat/thin (*needle*)	pointed objects – flat/thin (*shovel*)
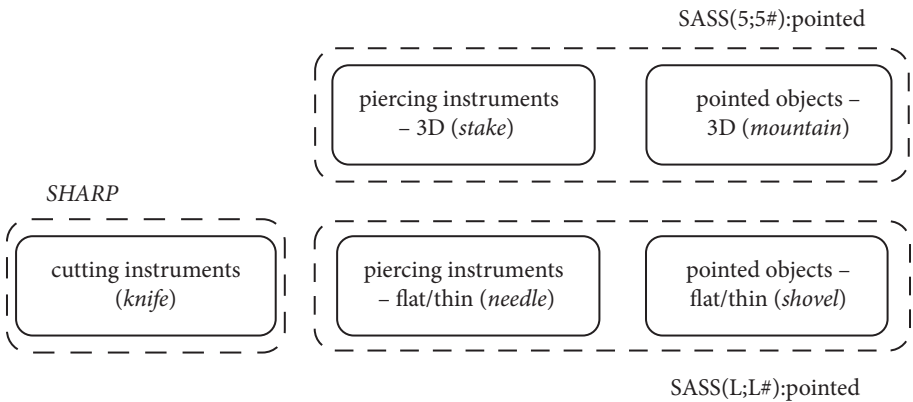
SASS(L;L#):pointed

Figure 16. RSL semantic domain SHARP

introduced in Sandler 1989). The two bottom frames are covered by the lexeme SASS(L;L#):pointed, where (L;L#) represents the ☝ -> ✍ sequence. This lexicalization strategy has not been attested in any spoken language investigated to date (see Chapter 2 in this volume)

RSL signs for size qualities can be analyzed in the same way. The domains THICK-WIDE and THIN-NARROW have parallel semantic structures consisting of five frames: 'thick/thin layers' (*thick/thin blanket, thick/thin sheet of paper*), 'thick/thin pivots' (*thick/thin column, thick/thin stick*), 'wide/narrow stripes' (*wide/narrow ribbon*), 'wide/narrow roads' (*wide/narrow road*) and 'wide/narrow holes' (*wide/narrow burrow, wide/narrow tunnel*), as in Figure 17 (see Chapter 5 of this volume).

layers *blanket, sheet of dust*	pivots *column, stick*	stripes *ribbon, plank*	roads *road, bridge*	holes *burrow, tunnel*
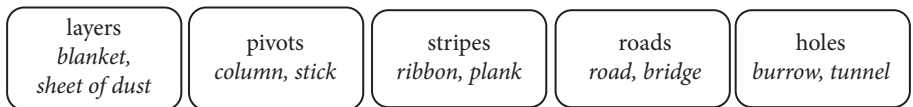
Figure 17. Semantic domains THICK-WIDE and THIN-NARROW

In RSL, different topological classes of objects are marked with different handshapes; see Table 2. The same handshapes are used for both 'thick-wide' and 'thin-narrow' meanings.

Table 2. Hand configuration in the domains THICK-WIDE and THIN-NARROW

Topological class	Hand configuration
layers	✍
pivots	✍;☝; ✊
stripes	✍
roads	✍ ; ☝
holes	☝

The remaining manual components in these signs are non-lexical and express the same meanings as corresponding structural elements of the non-core signs SASS(5;5#):pointed and SASS(L;L#):pointed, i.e. the hand orientation represents the orientation of the object, the localization component shows the object's position in space, and the movement characteristics describe its length and straightness/bending. The meaning of 'large' versus 'small size' is expressed by the non-manual articulation. 'Big size' is marked with the mouth gestures (MG) /af/, /o/, puffed cheeks, while the antonymous meaning is expressed with /peeh/, /pah/, /pl/, /pooh/, /mmm/. These mouth gestures are not limited to the semantic domains in question, but denote high versus low intensity in general. They also accompany signs expressing the qualities LONG/SHORT, TALL/LOW, BIG/SMALL. This distribution of chunks of meaning across the sign components allows the same manual sign to express antonymous qualities, as in Figure 18.

a. SASS(C)-thick (*plank*), MG /o/   **b.** SASS(C)-thin (*plank*), MG /peeh/

Figure 18. Antonymous meanings expressed by the same manual sign in RSL

In both (18a) and (18b), the two hands in the configuration 🤏 are moving in opposite directions in front of the signer. The handshape here denotes the topological class 'layer', as the signs describe planks. However, while the sign in (18a) expresses the meaning 'thick plank', the sign in (18b) describes a thin plank. The only formal feature these two signs differ in is the mouth articulation: the manual sign in (18a) is accompanied by the silent articulation /o/, which denotes big sizes, and the manual sign in (18b) is produced simultaneously with the mouth gesture /peeh/, which expresses small sizes.

Figure 19 shows the semantic map of the qualities THICK-WIDE and THIN-NARROW in RSL. For the sake of convenience, different lexemes are represented with different handshapes. For example, instead of "SASS(C)-thick/SASS(C)-thin" I use the picture 🤏; instead of "SASS(F)-thick/SASS(F)-thin", I use the picture 🤏; and so on. Different dash patterns stand for different non-core signs used in this zone.

The distribution of lexemes across the semantic space of THICK-WIDE and THIN-NARROW seen in Figure 19 is not attested in any other language analyzed so far. At the same time, the clear division between the 'thick'/'thin' and the 'wide'/'narrow'
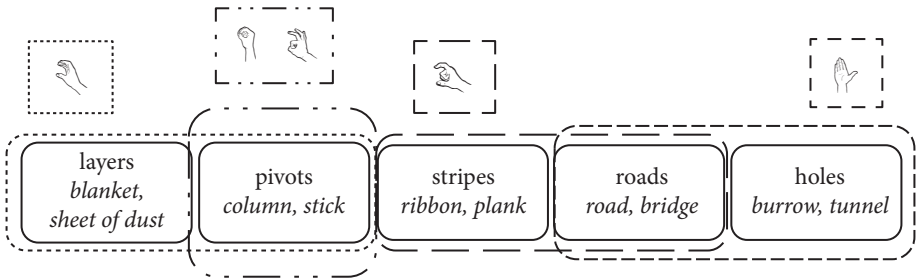
Figure 19. Semantic domains THICK-WIDE and THIN-NARROW in RSL

zones (i.e. the two left versus the three right frames, respectively), which is found in RSL, and the lack of a single SASS that would cover all the frames from both domains, is quite widespread typologically. For example, many European and a number of non-European languages (such as Hebrew, Udmurt, Armenian, Ingush) use a binary lexicalization strategy with one pair of words covering 'thick'/'thin' situations, and another pair covering 'wide'/'narrow' ones (Chapter 5 in this volume).

5. Discussion

This paper uses RSL qualitative domains to explore the possibility of a signed vs. spoken language lexicon comparison. Such a comparison becomes possible if we adopt the analysis according to which non-core signs contain two parts, gestural and lexical. This analysis allows for a direct application of the MLexT methodology in sign language lexicon research. As a result, we can discover RSL lexicalization strategies in different semantic domains. Some of these strategies have already been attested in spoken languages. In these domains, RSL data fits perfectly into already existing typological generalizations. In the domains where RSL shows unique lexicalization patterns, they still can be accounted for in the MLexT framework. Further research is necessary to clarify the nature of the signed versus spoken language lexicalization differences.

However, a point of significant divergence between Russian Sign Language and spoken language data is the absolute lack of semantic shifts in the former. Lexemes denoting different qualities in spoken languages often have not only literal, but also figurative meanings (*sharp knife – sharp image; high cliffs – high taxes*, to name a couple). The MLexT methodology, based on the study of linguistic contexts, can be applied to the analysis of figurative meanings to outline the range of metaphors that words from different semantic domains can have. For example, common semantic extensions of adjectives belonging to the semantic domain SHARP are 'intense pain',

'good sight/hearing', 'harsh wind', and some others (see Chapter 2 of this volume). However, this is not the case with RSL. All RSL qualitative signs analyzed in this study have only literal meanings and cannot be used metaphorically. This does not mean that sign languages do not have metaphors at all. For example, Taub (2004); Wilcox (2000); Meir (2010), and Brennan (2005) describe and analyze a large number of metaphors in American, British and Israeli sign languages. But as Taub (2004) points out, sign language metaphors are for the most part different from those in spoken language. In sign languages, it is rare for one lexeme to have both literal and figurative uses. Much more often a sign changes its form when used metaphorically, or a metaphoric sign does not express a literal meaning but is based on an iconic image instead (the latter is accounted for in the "analogue building model" of metaphorical iconicity, Taub 2004). However, the very possibility for comparison of sign and spoken language metaphors is still an issue open to debate, and more data (including from RSL, which is largely understudied in this respect) are necessary to fully address this issue.

Acknowledgements

References

Bellugi, U. & Klima, E. S. 1979. Language: perspectives from another modality. In *Brain and Mind* [Ciba Foundation Series 69], 99–117. Amsterdam: Excerpta Medica.

Benedicto, E. & Brentari, D. 2004. Where did all the arguments go? Argument-changing properties of Classifiers in ASL. *Natural Language and Linguistic Theory 22*(4): 743–810. https://doi.org/10.1007/s11049-003-4698-2

Berlin, B. & Kay, P. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley, California: University of California Press.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F. & Christiansen, M. H. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America* 113(39): 10818–10823. https://doi.org/10.1073/pnas.1605782113

Boyes-Braem, P. & Sutton-Spence, R. (eds). 2001. *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. Hamburg: Signum Verlag.

Brennan, M. 1990. *Word Formation in British Sign Language*. Stockholm: University of Stockholm Press.

Brennan, M. 2005. Conjoining Word and Image in British Sign Language (BSL): An Exploration of Metaphorical Signs in BSL. *Sign Language Studies 5*(3): 360–382. https://doi.org/10.1353/sls.2005.0007

Burkova, S. & Kimmelman, V. (eds). 2019. *Vvedenie v linguistiku žestovyx jazykov: Russkiy žestovyj jazyk* [Introduction to Sign Language Linguistics: Russian Sign Language]. Novosibirsk: NSTU

Cogill-Koez, D. 2000a. Signed language classifier predicates: Linguistic structures or schematic visual representation? *Sign Language & Linguistics* 3(2): 153–207. https://doi.org/10.1075/sll.3.2.03cog

Cogill-Koez, D. 2000b. A model of signed language 'classifier predicates' as templated visual representation. *Sign Language & Linguistics* 3(2): 209–236. https://doi.org/10.1075/sll.3.2.04cog

Emmorey, K. & Herzig, M. 2003. Categorical Versus Gradient Properties of Classifier Constructions in ASL. In *Perspectives on Classifier Constructions in Sign Languages*, K. Emmorey (ed), 221–247. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607447

Grose, D. 2012. Lexical semantics: Semantic fields and lexical aspect. In *Sign Language: an International Handbook*, R. Pfau, M. Steinbach & B. Woll (eds), 432–462. Berlin: De Gruyter Mouton.

Hanke, T. 2004. HamNoSys – representing sign language data in language resources and language processing contexts. In *Workshop Proceedings : Representation and Processing of Sign Languages*, O. Streiter & C. Vettori (eds), 1–6. Paris: ELRA.

Johnston, T. & Schembri, A. 2007. *Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511607479

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Liddell, S. K. 1980. *American Sign Language Syntax*. The Hague: Mouton. https://doi.org/10.1515/9783112418260

Liddell, S. K. 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511615054

Meir, I. 2010. Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language* 86(4): 865–896. https://doi.org/10.1353/lan.2010.0044

Perlman, M., Little, H., Thompson, B. & Thompson, R. L. 2018. Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology* 9: 132–147. https://doi.org/10.3389/fpsyg.2018.01433

Perniss, P., Thompson, R. L. & Vigliocco, G. 2010. Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1: 1–15. https://doi.org/10.3389/fpsyg.2010.00227

Pfau, R. & Steinbach, M. 2006. Pluralization in sign and speech: Across-modal typological study. *Linguistic Typology* 10: 135–182. https://doi.org/10.1515/LINGTY.2006.006

Pietrandrea, P. 2002. Iconicity and arbitrariness in Italian sign language. *Sign Language Studies* 2(3): 296–321. https://doi.org/10.1353/sls.2002.0012

Prillwitz, S., Leven, R., Zienert, H., Hanke, Th. & Henning, J. 1987. *HamNoSys. Hamburg Notation System for Sign Languages. An Introduction*. Hamburg: Zentrum für Deutsche Gebärdensprache.

Rakhilina, E. & Reznikova, T. 2016. A Frame-based methodology for lexical typology. In *The Lexical Typology of Semantic Shifts*, P. Juvonen & M. Koptjevskaja-Tamm (eds), 95–130. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110377675-004

Sagara, K. & Zeshan, U. 2016. Semantic fields in sign languages: A comparative typological study. In *Semantic Fields in Sign Languages: Colour, Kinship and Quantification* [Sign language typology 6], U. Zeshan & K. Sagara (eds), 3–41. Boston: De Gruyter Mouton.

https://doi.org/10.1515/9781501503429-001

Sandler, W. 1989. *Phonological Representation of the Sign Linearity and Nonlinearity in American Sign Language*. Dordrecht: Foris. https://doi.org/10.1515/9783110250473

Sandler, W. & Lillo-Martin, D. 2001. Natural Sign Languages. In *Handbook of Linguistics*, M. Aronoff & J. Rees-Miller (eds.), 533–562. Malden, MA: Blackwell.

Sandler, W. & Lillo-Martin, D. 2006. *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139163910

Stokoe, W. 1960. *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf* [Studies in Linguistics: Occasional Paper No. 8]. Buffalo, NY: University of Buffalo.

Supalla, T. R. 1986. The classifier system in American Sign Language. In *Noun Classes and Categorization*, C. Craig (ed), 181–214. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.7.13sup

Taub, S. F. 2004. *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press.

de Vos, C. & Pfau, R. 2015. Sign Language Typology: The Contribution of Rural Sign Languages. *Annual Review of Linguistics* 1: 265–288. https://doi.org/10.1146/annurev-linguist-030514-124958

Wilcox, P. P. 2000. *Metaphor in American Sign Language*. Washington, DC: Gallaudet University Press.

Wilkinson, E. 2009. Typology of Signed Languages: Differentiation through Kinship Terminology. PhD thesis, University of New Mexico.

Willbur, R. 2013. The point of agreement: Changing how we think about sign language, gesture, and agreement. *Sign Language & Linguistics* 16(2): 221–258. https://doi.org/10.1075/sll.16.2.05wil

Woodward, J. 1978. All in the family: Kinship lexicalization across sign languages. *Sign Language Studies* 19: 121–138. https://doi.org/10.1353/sls.1978.0016

Zeshan, U. (ed). 2006. *Interrogative and Negative Constructions in Sign Languages*. Nijmegen: Ishara Press. https://doi.org/10.26530/OAPEN_453832

Zeshan, U., Delgado, C. E. E., Dikyuva, H., Panda, S. & de Vos, C. 2013. Cardinal numerals in rural sign languages: Approaching cross-modal typology. *Linguistic typology* 17: 357–396.

Zwitserlood, I. 2003. *Classifying Hand Configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Utrecht: LOT.

CHAPTER 11

Constructing a typological questionnaire with distributional semantic models

Daria Ryzhova and Denis Paperno

HSE University, Moscow / Utrecht University

The paper presents a methodology for automatic construction of lexical typological questionnaires for qualitative semantic domains (e.g. SHARP, STRAIGHT, THICK, or SMOOTH). Our algorithm is based on data from a monolingual corpus; it constructs a list of collocations for the corresponding lexemes, computes a vector representation for every collocation, clusters the vector space into semantically homogeneous groups and extracts the three central elements from every cluster. We compare the resulting questionnaires against test data from the semantic domains that are already well studied manually. The algorithm demonstrates high quality results and can be used in the practice of lexical typological research.

Keywords: lexical typology, questionnaire, distributional semantic models, physical qualities

1. Introduction

The present volume reports the results of typological studies of several semantic domains representing basic qualitative features: WET, EMPTY, FULL, SHARP, BLUNT, THICK, THIN, and a few others. To perform an analysis of these domains, all the contributors followed the same basic methodology – the so-called frame-based approach to lexical typology (Rakhilina & Reznikova 2016). This chapter stands out from the rest of this volume as it focuses on the method itself rather than on introducing new data.

The previous chapters demonstrate that the frame-based approach appears promising and fruitful. The results and the data presented in this volume could be used for a variety of purposes encompassing further theoretical research as well as diverse practical applications. Nevertheless, this methodology has an evident disadvantage: it is very time- and labor-consuming. The projects presented in this book

cover only a tiny portion of the vocabulary; yet it took the researchers several years to complete them. If we proceed along these lines, then the goal of examining a significant part of the lexicon is likely to be virtually unattainable. The introduction of new, automatic methods of data processing could help overcome these problems, thus opening up new vistas for lexical typology.

When developing automated methods of data collection, a researcher may opt to rely on manually conducted case studies which adhere to the frame-based approach to lexical typology; they are extremely fine-grained and result in trustworthy datasets of good quality. The Moscow Database of Qualitative features, accumulating the data collected in the course of the research reported in this volume, can serve as a reliable source of benchmarks for computational models (see Ryzhova et al. 2016 for more details). The present chapter aims to evaluate the possibility of automating a part of this methodological pipeline, namely preliminary data selection and analysis. We are checking whether output questionnaires that our automated method produces are comparable to those obtained manually. Alternatively, the semantic regularities and oppositions that researchers reveal by hand may prove so subtle that they may escape detection in computational models. To explore these questions, the chapter describes an algorithm for automating the construction of a questionnaire, which comprises the first stage of any lexical typological research project, and evaluates its performance.

In Section 2, we briefly discuss the existing methodologies of questionnaire construction. Section 3 analyzes the problem of designing a questionnaire within the frame-based approach to lexical typology, including addressing the purposes served by a questionnaire, manual preparation of a questionnaire, and the processes that are to be automated. Sections 4.1 and 4.2 present the two main stages of the algorithm we suggest (i.e. collecting the list of possible contexts, and clustering the contexts). In Section 5, we evaluate our method against the data reported in studies on several semantic fields (SHARP, SMOOTH, STRAIGHT, and THICK). Finally, in the last part of the paper we discuss the results and sketch possible directions for expanding the algorithm to semantic fields other than qualitative features.

2.   Previous research

Linguistic typology aims to reveal constraints and regularities across the striking diversity of human languages. While grammatical typology focuses on expressions of grammatical meanings, studies in lexical typology address the differences between words, i.e. between so-called translational equivalents.

A questionnaire is an essential and sometimes the only available tool in typological research. In lexical typology, its role is even greater than in grammatical

typology for a number of reasons: there is normally little or no data on the meanings of lexical items in reference grammars; the information given in dictionaries is usually limited; and existing text corpora are often too small for the analysis of the lexicon. Despite the unquestionable importance of questionnaires, to the best of our knowledge, no explicit and principled methodology for their construction has been proposed thus far.

One exception (though limited to a specific domain) is the psycholinguistic approach in which the design of a questionnaire is predefined by a set of all the possible parameter values and their combinations. Perhaps the most famous example is the research on basic color terms across languages started by Berlin and Kay (1969) who constructed their set of stimuli based on the Munsell color system.

In contrast, data-driven approaches, such as the frame-based approach to lexical typology on which we rely here, strive to analyze textual data produced in natural situations; therefore, they do not enforce any hard and fast rules for designing lists of questions to be used in native speaker surveys. Usually, the starting point of a lexical typological project is a corpus of a well-resourced language. The researcher extracts from the corpus patterns of usage of the target items, and uses these patterns to build questionnaires for low-resourced languages.

A natural development of this methodology is to conduct research on the basis of parallel corpora – collections of aligned translations of the same texts into several languages. In this case, the list of contexts where the target items can potentially occur serves as an analogue of a "questionnaire" that is already pre-filled with data from multiple languages (cf. Dahl 2007; Wälchli & Cysouw 2012). The data collected in this fashion is as objective as possible, and the result is not affected by the possible initial bias of the researcher. Despite the advantages of this method, the sizes of existing parallel corpora render the method hardly operable for typological research on most lexical domains, since lexical items occur in texts much more rarely than grammatical markers. Furthermore, the lexical interference effects that arise in translated texts may affect the quality of parallel data.

The algorithm we suggest effectively overcomes the major problems listed above. Firstly, it follows a strict procedure of data collection and analysis which cannot be affected by a researcher's expectations or potential theoretical biases. Secondly, it is fully automatic and thus considerably saves a researcher's time and effort. Finally, it allows constructing a basis for a typological questionnaire relying on monolingual text corpora of a one single language. If one language suffices, one can choose as the starting point a sizable representative corpus for a high resource that covers different usages of diverse lexical items including rare ones.

3.   Typological questionnaires in the frame-based approach

There are several approaches to lexical typology which are guided by different research objectives and theoretical grounds (see the recent overview in Koptjevskaja-Tamm et al. 2016). The frame-based approach taken in the present volume rests on the assumption that semantic fields are sets of minimal lexical meanings (frames), and every lexeme from any language covers one of the subsets of the frames. For example, the domain of EMPTY comprises two main frames: 'hollow (being of a special shape: having a space inside)' and 'empty (having none of the usual or appropriate contents)'.[1] While some languages employ two distinct lexical terms to denote these meanings (cf. English *hollow* and *empty*, Serbian *šupalj* and *prazan*, Spanish *hueco* and *vacío*), other languages feature only one lexeme covering both of these frames (cf. the Khanty adjective *tăľ*).[2] Thus, the frame structure of a field serves as the basis for comparison between translational equivalents across languages.

Different frames are associated with different types of contexts. Returning to the previous example, the words with the meaning of 'hollow' usually modify names of geometrical shapes (*cylinder, sphere*) or nouns denoting bones and tree trunks, while the 'empty' lexemes typically combine with the names of containers (*box, cup, bag*, etc.). This implies that, in order to reveal the frame structure of a field as well as to establish the differences between lexemes within one language and across multiple languages, it is fruitful to scrutinize the distribution of the words from the domain in question.

The primary tool for addressing this task is a typological questionnaire. Most languages of the world and, consequently, most languages in appropriately selected typological samples, lack balanced and representative text corpora; therefore, native speaker surveys become the only practical method of data collection. This entails the following fundamental requirement: a questionnaire for a typological study of a certain semantic field should be able to foresee all the potential types of situations that can be described by the relevant lexemes in any language.

Traditionally within the frame-based approach, a preliminary version of a questionnaire is constructed manually on the basis of the researchers' native language – in the case of the studies reported in this volume, Russian (we will refer to it as *pivot language* below). The starting point is to pick the Russian lexemes that we believe to form the semantic field under investigation. Then we turn to the data of a reference corpus (in our case, the Russian National Corpus) to analyze the

---

1.   The definition is cited from http://www.dictionary.com/browse/empty.

2.   See Chapter 4 in this volume for further details.

distribution of the selected lexemes, i.e. the contexts in which they occur and how these contexts can be divided into more or less homogenous groups. In other words, we define the scope of usage of the lexemes in question and attempt to reveal the frame structure of the field.

Some oppositions within the semantic field in question – let us call them *surface oppositions* – are marked lexically in Russian and thus can be exhaustively elicited from the pivot language. For example, the field SMOOTH (see Chapter 6 in this volume) manifests three main groups of frames:

1. Objects that are visually even (like an even road or a terrain that is not hilly);
2. Objects that are smooth to the touch (like a child's skin);
3. Slippery objects that have no surface imperfections (but this property is negatively evaluated because it hinders the normal functioning of the object, for example, slippery floor).

In some languages, these three frames correspond to different lexemes, while other languages show two patterns of their colexification. First, one and the same lexeme can be used to describe visually even and smooth objects (frames 1 and 2), as opposed to slippery ones; second, in some languages, frames 2 and 3 corresponding to smoothness and slipperiness of an object are colexified, while frame 1 (a visual characteristic of an object) is lexically opposed to them.

In Russian, three adjectives cover this field, which exactly match the types of situations listed above: *rovnyj* 'even' covers the first situation, *gladkij* 'smooth' covers the second one, and *skol'zkij* 'slippery' the third one. Therefore, all the oppositions within the field are overtly marked with different lexemes in Russian, hence in this case Russian data is quite convenient to start with.

However, there may be cases where the pivot language features a lexeme which indiscriminately covers several situations from a semantic field, while some languages have several words to describe them. For example, such *hidden oppositions* in Russian are revealed when we look at the Izhma Komi lexemes *vol'k* and *vol'sjoony*, which divide the "slippery" situations into the two frames: the adjective *vol'k* describes bearing surfaces that are slippery (road, floor, etc.), but it does not apply to objects that slip out of one's hands (soap, fish, stick, and so on). The latter type of situation is covered by a different lexeme with the same root, *vol'sjoony*. In other words, objects that slip out of one's hands are lexically differentiated from surfaces that are difficult to walk on. In contrast, Russian uses a single adjective, *skol'zkij*, to describe both types of situations.

Revealing hidden regularities is quite a demanding task since, in these cases, the pivot language lacks overt lexical markers of the oppositions that can appear relevant in a typological study. When constructing a questionnaire manually, a

researcher usually relies on his/her own experience and semantic intuition in the pivot language, formulating and then verifying or falsifying hypotheses about possible oppositions within the field with data from other languages.

Automated methods of questionnaire construction can be useful for several reasons. First, the researcher will no longer have to manually analyze all the examples of a word's usage in a corpus, which would save substantial amounts of time and effort. Second, results will be more objective and independent of possible biases of the researcher. Finally, automation is expected to allow the researcher to quickly produce good quality preliminary questionnaires on the basis of only one language, thus further reducing the time and effort required to complete a lexical typological project.

4.   The algorithm for automatic questionnaire construction

A lexical typological questionnaire is a list of linguistic examples to be translated into the languages of the typological sample. Each example contains a lexical item of interest and a characteristic context which is representative of a frame, i.e. what we hypothesize is a cross-linguistically stable type of usage of the lexical item. The examples in a questionnaire contain as their basis certain minimal contexts that are "diagnostic" for revealing hidden semantic contrasts. The sketches on several qualitative semantic fields in the present volume (see also our examples from the fields EMPTY and SMOOTH in the previous section) demonstrate that "diagnostic" contexts for qualitative features are generally quite short and can be presented by a quality name and the name of an object to which the quality can be assigned, cf. a '(slippery) floor' or a '(smooth) skin'. These short basic diagnostic contexts serve as a common ground for cross-linguistic analysis, but in native speaker surveys, every item of a questionnaire is extended to a full sentence in order to provide a consultant with a natural broader context. In this paper, we suggest an algorithm that automatically construct a basis of a questionnaire, i.e., in case of the qualitative features, a list of noun phrases containing a modifier and the name of a modified object.

The problem of constructing such a questionnaire from monolingual (Russian) data boils down to two tasks:

1.   Collecting a list of contexts (or collocations) in which the target qualitative words (expressed with adjectives in Russian) occur;
2.   Dividing the contexts into groups (i.e. frames) and selecting a few examples of every frame.

Let us consider these tasks in succession.

4.1   Collecting a list of collocations

The first step of the algorithm is to collect a list of collocations for every Russian adjective in question.[3] Since we aim to construct a basis of a questionnaire consisting of minimal diagnostic contexts, it is sufficient in most cases to consider only the subject of the property expressed by the adjective. Such subjects are expressed by the noun which is modified by the adjective and which occurs next to it in the attributive construction (in Russian, the modifier usually immediately precedes the modified noun). This means that the list of possible contexts for a target lexeme consists of two-word noun phrases, bigrams of the type "adjective + noun".

To collect a list of bigrams with target adjectives, we used the main subcorpus of the Russian National Corpus (RNC); it is well balanced and, hence, provides more representative data than, for example, the Google bigrams collection.[4] Having access to a full machine-readable version of the main RNC subcorpus, we collected pairs of words consisting of the adjective in question and a noun located to its right. In contrast to most existing off-the-shelf bigram collections, such as Google N-grams cited above, we retrieve lemmas instead of word forms (such as *dom* 'house' and all its morphological forms *dom-a* 'house-GEN', *dom-u* 'house-DAT' etc. instead of the letter sequence *dom* only). Then, we eliminate from the resulting list of collocations all the bigrams that occur less than 10 times in our corpus, in order to avoid rare occasional word usages. This method produces a representative sample of noun phrases of the form "adjective + noun" and ensures that we collect a sufficient number of examples to illustrate each of the prospective frames of the semantic field under investigation.

4.2   Dividing the contexts into frames

Dividing a list of collocations into groups is a classical clustering task. In order to implement cluster analysis, we are to establish the basis for the comparison of bigrams. It is the semantic grounds that are crucial for the purposes of our analysis, as we want to obtain groups of collocations that are semantically close to each other, denoting typologically similar situations. We consider Distributional Semantic Models, as described in Section 4.2.1, particularly suitable for our task because they are designed to quantify semantic similarity between linguistic expressions.

---

3.   In this paper, we leave aside the problem of defining the set of Russian adjectives that constitute a given field. To the best of our knowledge, the notion of a semantic field has no strict definition, so the choice of lexemes is to a certain extent arbitrary.

4.   See http://storage.googleapis.com/books/ngrams/books/datasetsv2.html, the most popular resource for n-grams in English, Russian, and several other languages.

4.2.1   *Distributional semantic models*

Distributional semantics is based on the distributional hypothesis, according to which linguistic items with (dis)similar distributions tend to have (dis)similar meanings (see Sahlgren 2008 for a discussion). In the Distributional Semantic Models (DSM) framework (see Baroni et al. 2014), the meaning of a lexical item is identified as the set of its co-occurrences in a training corpus, and is represented as a vector where the context words play the role of dimensions. In the simplest case, the value of a dimension is the number of times the target lexeme occurs near the word serving as the dimension. The notion of "near the word" (or "in the context of the word") is defined differently in each individual distributional semantic model: some models operate on a window of a fixed size from the target word, while others look at certain syntactic distances in terms of dependency grammar. For example, the ±3 word window means that the target lexeme should occur not further than 3 items to the left or to the right of the context word.

The frame approach to lexical typology parallels the DSM methodology in its reliance on context (cf. also research on English temperature terms in typological perspective based on the distributional semantic modelling in Koptjevskaja-Tamm & Sahlgren 2014). While both paradigms rely on the distributional properties of a lexical item in question, the DSM framework has two additional properties arising from its computational nature:

1. It considers a far larger set of contexts than what is feasible to analyze manually;
2. It represents the meaning of a lexeme as a mathematical object (a vector) which one can operate with (for example, these objects can be compared to each other and clustered automatically).

In our work, we construct a vector representation for every noun phrase from the list collected in the previous stage. The basic parameters of our model are as follows: as the training text set, we use the main subcorpus of the Russian National Corpus (about 220m tokens); every vector is derived from the co-occurrence frequency of the 10,000 most frequent content words in our corpus (i.e. nouns, verbs, adjectives, and adverbs) within the ±5word window from the target lexical item.[5]

Note that the units in the resulting list of bigrams are not single words but two-word noun phrases. There are two possible ways to compute the vector

---

5. The cooccurrence counts are then weighted using PPMI (Positive Point-wise Mutual Information) and passed through Singular Value Decomposition (SVD), which reduces the number of vector dimensions from 10,000 to 300. We used the implementations of these transformation schemes from the DISSECT toolkit (http://clic.cimec.unitn.it/composes/toolkit/index.html, see also Dinu et al. 2013). The above parameter values are standard in DSM and have shown consistently good performance in diverse empirical evaluations. See Ryzhova et al. (2016) for an extensive discussion of these parameters in relation to lexical typology.

representation for a multiword item of this kind. First, a noun phrase can be treated as an indivisible item. In this case, we analyze the distribution of the phrase as a whole, as if it were a single token. To compute the phrase's vector under this scenario, we take only the sentences where these two lexemes occur together, the adjective before the noun, without any intervening words. Of course, this solution leads to a dramatic shrinkage in the size of the data, as collocations are much less frequent than their constituent lexemes. A plausible vector representation of a multiword expression can therefore be computed only on the basis of a very large corpus, and only for high-frequency phrases.

Another option is to compose the vector representation of a two-word collocation from the corresponding distributional vectors of its parts. This method might seem less straightforward, but it is more economical and efficient, and it yields adequate results in practice. According to this *compositional* method, vector representations for individual lexemes are computed first, and then one of the composition models is applied to derive a representation for the entire phrase. Several composition models have been proposed for this purpose (such as additive, multiplicative, lexical function, etc.).[6] In our task of automatic questionnaire construction we applied one of the simplest schemes – the additive model – because it had steadily demonstrated good results in previous research (consider Blacoe and Lapata 2012; Ryzhova et al. 2016, among the others). This method consists in summing up the vectors of the constituent lexemes of a phrase, i.e. we perform pairwise summation of the values of each dimension.

4.2.2   *The clustering algorithm*

Every context (i.e., a two-word noun phrase) is already represented as a vector (i.e., a sequence of numbers) and can be treated as a point in a high-dimensional space (see Section 4.2.1). A clustering algorithm finds the points that are close to each other in the vector space and combine them in a cluster. Since a lexical typologist at the start of a project is not aware of how many frames constitute the semantic field under investigation, we applied the hierarchical clustering algorithm (Johnson 1966) that does not require the number of clusters to be specified in advance.[7] Every cluster is an area in a vector space, with points representing noun phrases located inside it. For each area, the central point called *centroid* can be computed. In most cases, it is just a

---

6.   See Mitchell and Lapata's (2010) groundbreaking paper for discussion. All the popular models are present in the DISSECT toolkit that we used for our experiments.

7.   The algorithm was implemented by means of the SciPy module (http://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html#module-scipy.cluster.hierarchy); it was run with the default criterion parameter ('inconsistent'), the cosine similarity metric and the clustering threshold of 1. The parameter settings were tuned for best performance; however, these details are beyond the scope of the present paper.

mathematical object – a point with some coordinates in the same high-dimensional space, and it does not correspond to any of the noun phrases constituting the cluster. A centroid can be considered as an idealized representation of a prototype meaning of a cluster. Technically, the values of its dimensions are calculated as the mean of the corresponding values of all the vectors grouped into the cluster.

For example, the clustering algorithm puts into the same cluster the noun phrases *ostryj_zubec* 'sharp cog', *ostryj_veršina* 'sharp peak', *ostryj_konus* 'sharp cone', *ostryj_skala* 'sharp rock', *ostryj_vystup* 'sharp ledge' and others, some of which are more likely to denote sharp quasi-instruments (dangerous objects that can cause a damage, such as a rock or a ledge). Computing a centroid of this cluster and extracting three elements closest to it result in a more coherent picture: the core elements are *ostryj_zubec* 'sharp cog', *ostryj_skala* 'sharp rock', and *ostryj_vystup* 'sharp ledge', i.e. the cluster represents specific potentially dangerous quasi-instruments, while *ostryj_veršina* 'sharp peak' and *ostryj_konus* 'sharp cone' turn out to be peripheral, in part because they are less instrument-like and in part because of the ambiguity of the nouns *veršina* ('peak (of a mountain)' vs. 'vertex (of a polygon)') and *konus* ('cone' as a physical shape, e.g. of a tree, vs. as a geometric shape).

After we have clustered the list of contexts, we exclude from consideration clusters containing less than three elements, and for all the remaining clusters, we compute the centroids and extract the three core elements, i.e. those that are the closest to the centroid (as measured by the cosine similarity metric) from every cluster, as the basis of a questionnaire. The solution of extracting three core elements from every cluster is motivated by two reasons. First, it shortens the resulting questionnaire; a questionnaire is not practical for surveying native speakers if it is excessively long. Second, it allows us to filter out peripheral elements that are far from a cluster's prototype. In most cases, these are ambiguous noun phrases denoting different meanings depending on a broader context, such as the Russian equivalents of 'sharp cone' and 'sharp peak' above. To give another example, 'sharp beak' can be assigned to one of two different frames. Consider the following sentences:

(1) *Birds that hunt for fish in shallow water have long, **sharp beaks** that they use to <…> spear fish.*

(2) *Female hummingbirds use their long, tubular tongues and **sharp beaks** to feed their babies.*

The *sharp beak* from sentence (1) is certainly a piercing instrument, while the same noun phrase in Example (2) is more likely just to describe the shape of the object rather than a piercing tool since it does not perform any piercing, certainly not piercing the chicks to be fed. These situations are regarded as two different frames ('sharp piercing instrument' vs. 'object of a sharp form'),[8] because in some languages

---

8.   See Chapter 2 in the present volume for further details.

they are expressed by different lexical means. A core, prototypical exemplar of the first frame is an object like a spear or an arrow, while a sharp nose or the pointed toes of a shoe would belong to the second one, representing unambiguously the semantics of each frame. Objects like a sharp beak are ambiguous and hence peripheral for both frames since in some situations they are used as piercing tools, while in others they are represented as objects of a certain shape. In terms of a vector space, we expect that the noun phrases denoting such objects would occur in different types of contexts, and thus they should not appear near the centroid of any cluster, being located instead somewhere on the border between two groups of items, cf. a similar assumption in Dubossarsky et al. (2016) for ambiguous lexemes.

5. Evaluation

5.1 The metric

In order to evaluate the results obtained by the suggested method, we chose four semantic fields that had been previously investigated manually: SHARP (constituted by the Russian adjectives *ostryj* and *rezkij* 'sharp'), STRAIGHT (represented by *prjamoj* 'straight, direct'), SMOOTH (with the lexemes *gladkij* 'smooth', *skol'zkij* 'slippery', *rovnyj* 'even', and *ploskij* 'flat'), and THICK (*tolstyj* 'thick' and *širokij* 'wide, broad'). For details on the domains SHARP, SMOOTH and THICK, see the respective Chapters 2, 6 and 5 in the present volume; for the STRAIGHT domain, see Luchina, Reznikova, & Stenin (2013).

We asked experts on those semantic fields to mark up the lists of collocations that were obtained automatically at the first stage of the algorithm. In particular, the experts were instructed to indicate the frame to which every noun phrase belongs; assigning of more than one frame number was allowed. A fragment of an expert's markup of the list for the field STRAIGHT is presented in Table 1.

Table 1. STRAIGHT: An expert's markup. The numbers correspond to different frames. The pipe symbol "|" divides multiple frame values

Collocation	Frame(s)
prjamoj rjad 'straight row'	1
prjamaja linija 'straight line'	1
prjamoj udar 'straight blow' / 'head-on attack'	1|4
prjamoj dostup 'direct access'	6
prjamoj razgovor 'direct/frank conversation'	6|7
prjamoj ugroza 'direct threat'	7

The output of the algorithm was evaluated against the experts' markup. We computed recall (R, the number of frames presented in the final version of the automatically produced questionnaire, divided by the total number of frames present in the expert's markup) and precision (P, in our case, the clustering purity – the standard metric measuring to which extent automatically produced clusters are homogenous, i.e. contain elements of a single frame), and then we calculated a balanced F-score which incorporates both precision and recall and thus can be treated as an overall assessment of the resulting automatically constructed questionnaire (see Formula 1). All these metrics take values in the range from 0 to 1, with 0 indicating the worst and 1 indicating the best possible performance of an algorithm.

$$F = 2PR / (P+R)$$

Formula 1. F-score ("P" = Precision, "R" = Recall)

The elements that had been assigned to multiple frames by the experts were considered to belong to several groups of the manual markup at the same time. Thus, we treated the cluster consisting of *prjamoj rjad* 'straight row' (frame 1), *prjamaja linija* 'straight line' (frame 1), and *prjamoj udar* 'straight/direct attack' (frames 1 and 4) as a pure one yielding the maximum precision value of 1 in such cases.

Table 2. Evaluation ("R" = Recall, "P" = Precision, "N of clusters" = number of clusters)

	Algorithm output				Best random clusterings			
	R	P	F-score	N of clusters	R	P	F-score	N of clusters
SHARP	0.733	0.827	0.777	25	1	0.43	0.6	116
STRAIGHT	1	0.817	0.899	20	1	0.59	0.74	46
SMOOTH	0.8	0.675	0.732	38	0.999	0.5	0.669	109
THICK	1	0.884	0.938	23	1	0.53	0.69	21

The results in Table 2 show that the performance delivered by the algorithm was rather strong for all of the evaluated semantic fields (columns "Recall", "Precision" and "F-score" in the left part of the Table). We also check whether our clustering algorithm surpasses the quality of a random division of the collocations into groups (the right-hand part of the Table). For these purposes, we randomly split the lists of collocations derived from the corpus at the first stage of the algorithm for every field into clusters of three elements, varying the number of clusters from 1 to N/3, where N is the total number of noun phrases in the list of collocations for the field. We iterated this operation 100 times, i.e., for every field, we generated 100 questionnaires

containing only one cluster, 100 questionnaires with two clusters, etc. For each of the obtained questionnaires, we computed recall and precision, and averaged them across the clusterings of equal sizes (for example, the average recall for the random 20-clustering division for the field SHARP equals 0.698, and the average precision is 0.425). Based on these averaged values of the metrics, we computed the F-score according to Formula 1. The four rightmost columns of Table 2 present the highest scores and the number of clusters in the most successful random clusterings.

A closer look at the random splits reveals the following picture: within a semantic field, the F-score value rises with the increase of the total number of clusters (consider Figure 1). This is due to the recall that gradually goes up and at some point reaches the maximum (this point depends on the field, namely, on the total number of frames and on the number of noun phrases representing each frame). The precision (purity) value of random splits steadily hovers around chance (about 0.5), while the hierarchical clustering algorithm always demonstrates better clustering purity.
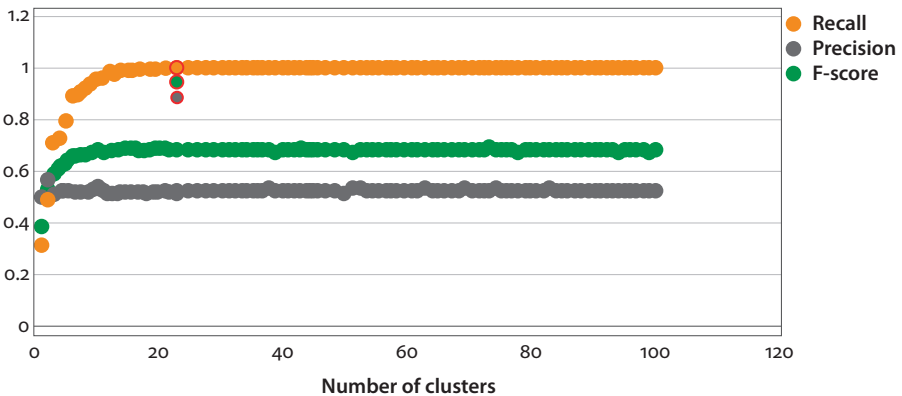
Figure 1. Random clustering of the field THICK. The dots with the red borders indicate the performance of our algorithm (top down: Recall, F-score, precision)

5.2   Qualitative analysis of the obtained clusterings

The algorithm presented in this paper demonstrates that automation of at least some stages of lexical typological research is not only possible, but also can be quite successful. Qualitative analysis of the best clusterings shows that the resulting questionnaires are informative, cf. a fragment of the clustering of the field STRAIGHT in Table 3.

Table 3. Fragment of the best clustering for the field STRAIGHT

Cluster 1	Cluster 2	Cluster 3	Cluster 4
prjamoj_stolb 'straight pole'	*prjamoj_učastie* 'direct participation'	*prjamoj_potomok* 'direct descendant'	*prjamoj_ neobxodimost'* 'urgent necessity'
prjamoj_dorožka 'straight path'	*prjamoj_podderžka* 'immediate support'	*prjamoj_ predšestvennik* 'direct predecessor'	*prjamoj_cel'* 'primary aim'
prjamoj_alleja 'straight avenue'	*prjamoj_rukovodstvo* 'immediate control'	*prjamoj_nasledie* 'direct heritage'	*prjamoj_zadača* 'immediate task'

Nevertheless, the quality of the algorithm's output depends on the dataset.

First, higher frequency adjectives are more likely to be clustered efficiently. However, even frequent words can be relatively rare in some of their usages. For example, only five noun phrases representing the frame 'intensive (of weather phenomena)' in the field SHARP occurred in the list of bigrams, which contains 351 phrases in total. This observation explains why any of the noun phrases illustrating the frame 'intensive (of weather phenomena)' do not appear in the final version of the questionnaire, lowering the recall value (see Table 2 above).

The second relevant property of the dataset is related to the first one: semantic fields with a relatively small number of frames are better clustered than more complex systems. The favorable scenario is realized in the case of the field STRAIGHT: according to the expert markup, it contains seven frames, and every frame is frequent enough to ensure high clustering quality.

Finally, the resulting clusterings reflect the taxonomic classification of the nouns that combine with the target adjectives. In most cases, it is precisely what we need, since adjectives in different senses typically collocate with different nouns. Consider, for example, a fragment of the best clustering for the field STRAIGHT in Table 3, where *dorožka* 'path' and *alleja* 'avenue' fall into one cluster, while *potomok* 'descendant' and *predšestvennik* 'predecessor' form another group. These two clusters correspond to two different senses of the adjective *prjamoj*: (1) having no bend or curve (cf. *straight path* and *straight avenue*); (2) related through parents, grandparents etc., and not through other relatives (cf. *direct descendant* and *direct predecessor*).

However, sometimes the semantic structure of a field is more complicated. For example, within the field SMOOTH, several adjectives can modify the same nouns, highlighting different aspects of an object: *skol'zkij **pol*** 'slippery floor' vs. *rovnyj **pol*** 'flat floor'; *gladkij **kamen'*** 'smooth stone' vs. *ploskij **kamen'*** 'flat stone'. Similarly, the noun phrases *gladkaja stena* 'smooth wall' and *ploskaja kryša* 'flat roof' quite undesirably end up in one and the same cluster. Separate clustering of every adjective's list of contexts results in a better performance in terms of the F-score in

such cases: for example, noun phrases with the adjective *gladkij* 'smooth' clustered independently from the other contexts demonstrate the 0.923 F-score (as opposed to 0.732 when the contexts of all the adjectives are clustered en masse, see Table 2).

Sometimes a clustering error arises in oppositions that rely on a topological classification of nouns rather than on a taxonomic one. In the field SHARP, instruments with a cutting edge (e.g. knife or sword) are cross-linguistically often described with a different adjective than instruments with a piercing point (such as spears and arrows). Since the taxonomic class ('instruments') is the same in both cases, and the difference consists only in the topological type of the salient segment (line vs. point), it is very problematic to uncover this opposition automatically. Our automatic clustering of the field SHARP produced a perfect cluster for cutting instruments: *ostryj nož* / *ostryj nožik* 'sharp knife' and *ostroe lezvie* 'sharp blade'. At the same time, it did not identify a cluster that would correspond to piercing instruments exclusively. The best candidate cluster contains the noun phrases *ostroe kop'ë* 'sharp spear', *ostryj meč* 'sharp sword' and *ostraja strela* 'sharp arrow'. However, this division reflects the taxonomic difference between household instruments as contrasted with weapons, rather than the topological opposition between 'line' and 'point'.

Interestingly, the dataset THICK demonstrates unexpectedly high quality of clustering. While the adjectives *tolstyj* 'thick' and *širokij* 'wide, broad' are highly frequent, and the number of frames that we investigated is quite small (we took only six frames into account), almost all the oppositions in this field are topological in nature: cf. 'thick pivots (tree, pencil, rope, etc.)' vs. 'thick layers (book, blanket, wall, etc.)' vs. 'wide stripes (ribbon, road, river, etc.)' vs. 'wide holes and tubes (window, door aperture, tunnel, etc.)' vs. 'wide surfaces (table, field, lake, etc.)'.[9] Nevertheless, the obtained clustering suggests that the topological classification correlates well with the taxonomic one. For example, streets and avenues are usually lexicalized as 'stripes', while skirts, dresses and trousers are lexicalized as 'tubes'. Nevertheless, there are some exceptions, and they cause mismatches: in our clustering, *tolstaja stena* 'thick wall' ('layer'), *tolstyj stolb* 'thick post' ('pivot'), an*d tolstoe brevn*o 'thick log' ('pivot') are placed into the same cluster.

---

9. In line with the terminology introduced in Chapter 5 of the present volume, we call 'pivots' elongated objects in which one dimension is considerably larger than the second and the third ones; 'layers' are flat objects with a rather small distance between their lower and upper surface; 'stripes' are flat elongated objects with one dimension considerably surpassing the second and the third ones; 'holes' are empty spaces in layer-like objects; 'tubes' are elongated empty spaces; 'surfaces' are spaces which are not elongated and which in general lack the third dimension.

The crucial role of taxonomic classification of the nouns collocating with the adjectives in question leads to the idea that an automatic clustering of the vector representations for the nouns (instead of the phrases "adjective + noun") would be sufficient for our purposes. This means that after constructing a list of "adjective + noun" collocations for every semantic field, we could compute vector representations for nouns extracted from these collocations (e.g. for *nož* 'knife' instead of *ostryj* 'sharp' + *nož* 'knife'; for *stolb* 'post' instead of *tolstyj* 'thick' + *stolb* 'post') and cluster them. The resulting groups of nouns could provide a key to distributional properties of the adjectives in question, e.g. 'the adjectives from the field SHARP can combine either with the names of weapons, or with the names of diseases etc.'. Indeed, our preliminary experiments tend to support this idea, showing that a cluster of a noun, though automatically assigned, is a good predictor of an adjective's meaning (see the results of the clustering of the corresponding noun-only vector spaces in Table 4).

Table 4. Evaluation of the noun-only model

	Recall	Precision	F-score
SHARP	0.7	0.88	0.78
STRAIGHT	0.947	0.611	0.743
SMOOTH	0.857	0.73	0.789
THICK	1	0.821	0.901

We hypothesize that these results are so comparable to those reported in Table 2 mostly due to the composition model we used: since it presupposed simple addition of the vectors of a noun and an adjective, it could not capture more complex semantic dependences between the parts of a noun phrase. This means that there are two possible avenues for further research on the suggested methodology:

1. On the one hand, relying on vector representations of just nouns instead of two-word noun phrases may be more parsimonious. But, this methodology should be used with caution: in some cases, it leads to significantly worse results (compare the results on the STRAIGHT domain in Tables 2 and 4).
2. On the other hand, applying more complex composition models may potentially lead to better performance of the algorithm.

6.   Discussion

The method proposed in this chapter consists of the four main steps:

1.   Collect a list of contexts in which the target words occur (our study is based on the data of the lemmatized main subcorpus of the RNC);
2.   Create a semantic vector space (with the Distributional Semantic Models techniques);
3.   Cluster the resulting space (with the hierarchical clustering algorithm);
4.   Extract three core elements from every cluster.

Each of these stages is completely automatic. The algorithm divides the contexts of usage of the adjectives from a given semantic field into homogenous groups with three exemplars each, revealing thus potential lexical oppositions within the field. These sets of clusters may serve as a basis for a typological questionnaire: the noun phrases that constitute them may be translated into the language under study and extended to full-sentenced contexts which, in their turn, may be used in various elicitation tasks (sentence translation, filling the gap etc.).

We applied the methodology to qualitative semantic domains. But the algorithm can be easily applied to the analysis of lexical fields other than qualitative features. Construction of a questionnaire for lexemes from other parts of speech or other types of semantic domains would require minimal adjustment of the methodology. The changes would have to be introduced only in the first step of the algorithm – collecting a list of collocations, – since different parts of speech possess different distributional properties.

For example, our preliminary experiments with verbs of oscillation show that a "diagnostic" context for intransitive oscillation verbs may consist of the verb's subject only. Since our corpus has no syntactic markup, we assume that the subject is in most cases expressed by the noun in nominative case occurring next to the verb in question. In Russian, the noun performing this role can occur either to the left or to the right of the verb. Consequently, the list of contexts should contain bigrams of two types: "noun.NOM + verb" (*derevo kačaetsja* 'a tree sways') and "verb + noun. NOM" (*kačaetsja derevo*, with the same meaning). This approach, which naturally extends from adjectives to intransitive verbs (such as *padat'* 'fall'), could be further applied to verbs that take more than one argument (such as *vytaskivat'* 'take something out of something'). This, however, might require more significant changes at the first stage of the algorithm, since the diagnostic contexts of such verbs could be more complex, involving the multiple arguments of the verb. We leave the exploration of these issues for future research.

Along with the type of diagnostic context, the threshold of the n-grams' frequency may also require modification. In the experiments reported in this chapter, we selected only the noun phrases that occurred not less than 10 times in our corpus. This value is optimal for analysis of lexemes like *ostryj* 'sharp' or *prjamoj* 'straight', which rank 1,452 and 892 correspondingly in the Russian frequency dictionary;[10] however, this threshold may not be as successful for analyzing less frequent words.

For instance, the adjective *prostornyj* 'spacious' (rank 4,421) in Russian can describe roomy spaces (*prostornyj zal* 'spacious/roomy hall') as well as loose clothes (*prostornaja rubaxa* 'loose shirt'). If we attempt to extract its list of collocations with a threshold of 10, we will retrieve the contexts for the first type of situation only. We will miss the phrases with nouns denoting clothes, as they occur less than 10 times in the RNC main subcorpus (*prostornaja rubaxa* 'loose shirt' – 9, *prostornyj pidžak* 'loose coat' – 8, *prostornoe plat'e* 'loose dress' – 6 times). As a consequence, the sample of collocations will not be representative.

Therefore, the threshold should be lowered when dealing with low-frequency lexical items. In our pilot experiments with verbs of oscillation (ranking between 4,802 and 14,321 in the Russian frequency dictionary) we applied the threshold of 3 occurrences and received adequate results: the Recall value of 0.882, the Precision of 0.762, and the F-score of 0.818.

Theoretically, one could try to apply this minimal threshold to any semantic field under investigation; however, such simplification would not be an optimal decision. In the case of higher-frequency items, such a low threshold dramatically inflates the lists of contexts, decelerating their processing. For example, the main RNC subcorpus contains 200 noun phrases with the adjective *ostryj* 'sharp' that occur 10 or more times; lowering the threshold to the value of 3 expands the list to the size of 541 items. It may be that for the top-frequency lexemes the optimal threshold would be even higher than 10 (up to 50 for adjectives like *xorošij* 'good'). We suppose that this value can be defined either in terms of a sufficient number of bigrams, or as a function of the frequency rank of the target term; this is an issue for future research.

Besides the diagnostic context and the frequency threshold, the rest of the algorithm can be left intact if we want to apply it to other semantic fields or parts of speech.

---

10. Prepared by O. Lyashevskaya and S. Sharov (http://dict.ruslang.ru/freq.php).

7.   Conclusion

To sum up, this paper has demonstrated a method for automatic questionnaire construction. The method proves capable of revealing most of the relevant oppositions within a given semantic domain using the data of only one language (including the oppositions that are not lexicalized in this language). We successfully tested the algorithm on four qualitative semantic fields. However, preliminary experiments show that it is also applicable to other types of semantic domains, reflected in non-adjective classes. The algorithm is fully functional and is ready to be employed in typological practice. Now that we can automatically compile a questionnaire, the next step in the direction of automating lexical typology is to devise a method that would automatically fill the questionnaire with data from diverse languages.

Acknowledgements

References

Baroni, M., Bernardi, R. & Zamparelli, R. 2014. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* 9: 241–346.

Berlin, B. & Kay, P. 1969. *Basic Colour Terms: their Universality and Evolution*. Berkeley, CA: University of California Press.

Blacoe, W. & Lapata, M. 2012. A comparison of vector-based representations for semantic composition. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 546–556. Jeju Island, Korea: Association for Computational Linguistics.

Dahl, Ö. 2007. From questionnaires to parallel corpora in typology. *STUF (Sprachtypologie und Universalienforschung)* 60(2): 172–181. https://doi.org/10.1524/stuf.2007.60.2.172

Dinu, G., Pham, N. & Baroni, M. 2013. DISSECT: DIStributional SEmantics Composition Toolkit. In *Proceedings of the System Demonstrations of ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*, 31–36. East Stroudsburg PA: ACL.

Dubossarsky, H., Weinshall, D. & Grossman, E. 2016. Verbs change more than nouns: a bottom-up computational approach to semantic change. *Lingue e linguaggio* 1: 7–28.

Johnson, S. C. 1966. Hierarchical clustering schemes. *Psychometrika* 32(2): 241–54. https://doi.org/10.1007/BF02289588

Koptjevskaja-Tamm, M., Rakhilina, E. V. & Vanhove, M. 2016. The semantics of lexical typology. In *The Routledge Handbook of Semantics*, N. Riemer (ed), 434–454. London, New York: Routledge.

Koptjevskaja-Tamm, M. & Sahlgren, M. 2014. Temperature in the Word Space: Sense exploration of temperature expressions using word-space modeling. In *Linguistic Variation in Text and Speech, within and across Languages*, B. Szmrecsanyi & B. Wälchli (eds), 231–267. Berlin/Boston: Mouton de Gruyter. https://doi.org/10.1515/9783110317558.231

Luchina, E., Reznikova, T. & Stenin, I. 2013. Atributivy kak istočnik grammatikalizacii: 'pryamoj' i 'rovnyj' v russkom, nemeckom i finskom jazykax [Attributives as a source for grammaticalization: 'straight' and 'even' in Russian, German, and Finnish]. In *Tipología Léxica*, R. Guzman Tirado & I. A. Votyakova (eds) 123–130. Granada: Jizo Ediciones.

Mitchell, J. & Lapata, M. 2010. Composition in distributional models of semantics. *Cognitive Science* 34(8): 1388–1429. https://doi.org/10.1111/j.1551-6709.2010.01106.x

Rakhilina, E. & Reznikova, T. 2016. A Frame-based methodology for lexical typology. In *Lexico-Typological Approaches to Semantic Shifts and Motivation Patterns in the Lexicon*, M. Koptjevskaja-Tamm & P. Juvonen (eds), 95–130. Berlin, Boston: Mouton De Gruyter. https://doi.org/10.1515/9783110377675-004

Ryzhova, D., Kyuseva, M. & Paperno, D. 2016. Typology of adjectives benchmark for compositional distributional models. In *Proceedings of the Language Resources and Evaluation Conference*, 1253–1257. Paris: European Language Resources Association (ELRA).

Sahlgren, M. 2008. The distributional hypothesis. *Italian Journal of Linguistics* 20: 33–53.

Wälchli, B. & Cysouw, M. 2012. Lexical typology through similarity semantics: toward a semantic map of motion verbs. In M. Koptjevskaja-Tamm & M. Vanhove (eds.), *New Directions in Lexical Typology. Linguistics (a special issue)* 50(3): 671–710. https://doi.org/10.1515/ling-2012-0021

Languages index

Subject index

What is it like? – This is often the first question we ask about any object, and it is typically answered with adjectives: *old, smooth, pointed, narrow,* etc. Characteristics of things around us is a fundamental aspect of how we conceptualize the physical world, regardless of when or where we live – and regardless of our language. Despite this, the vocabulary of physical qualities has received comparatively little attention in lexical typology: most research so far has focused on verbs and the actions they express.

This volume presents a lexico-typological study of several domains of physical qualities: 'sharp'/'blunt', 'wet', 'empty'/'full', 'old', as well as dimensions temperature and surface texture. It discusses several theoretical issues including intragenetic language sampling, the possibility of signed vs. spoken language comparison at the lexicon level, and the potential of applying computational models of distributional semantics to lexical typology.

The book will be of interest to linguists with a focus on typology, general and lexical semantics, to lexicographers, and to language students and teachers.

John Benjamins Publishing Company