

APPENDIX A: RESEARCH PROGRAM

1. *Specific aims*

1.1. The main issue dealt with in the proposed project is the conceptualisation of temperature in natural languages as reflected in their systems of central temperature terms, such as *hot*, *cold*, *to freeze*, etc. The project is envisaged as an INTEGRATED LEXICAL-TYOLOGICAL STUDY OF THE TEMPERATURE DOMAIN across a number of genetically, areally and structurally diverse languages with the aim of describing and accounting for the cross-linguistic variation within it from three different angles:

Lexicalization of temperature concepts, categorization within, or “carving up” the temperature domain: What temperature concepts are encoded as words across languages, what distinctions are made in the systems of temperature terms and what factors underlie them? Are there universal temperature concepts? Can temperature terms and temperature term systems completely free to vary across languages, or are there limits to this? How can the meanings of temperature terms be described (e.g., via reference to the objective temperature scale, to the human body and human perception or to typical entities, like fire or ice)?

Lexicon-grammar interaction within the temperature domain: How are temperature concepts lexicalized across languages in terms of word classes? What syntactic constructions are used for talking about temperature perception?

Semantic derivation and motivation (patterns of polysemy and semantic change) relevant for the temperature domain: What are the possible semantic extensions of the temperature meanings to other domains and how can these be related to their concrete meanings? Where from do the temperature terms come? How can the meaning of the temperature terms change within the temperature domain itself? What general metaphorical and metonymical models underlie the semantic evolution of the expressions related to the temperature domain?

In accordance with modern typological standards, the project aims at achieving the following results for each of the research topics: **description** of the attested phenomena, of the cross-linguistic variation within them and of the parameters behind it; **generalizations** about the attested cross-linguistic patterns and the limits to the cross-linguistic variation, and **explanations** for the attested cross-linguistic patterns and for their distribution across languages (linguistic, neurophysiological, cognitive and further biological, environmental, socio-cultural, areal and genetic factors).

1.2. On a more general level, the project is envisaged as a contribution to a wide range of THEORETICAL DISCUSSIONS focusing on meaning, language vs. cognition vs. culture, linguistic categorization, interaction between lexicon and grammar, relations between synchrony and diachrony, polysemy, etc.

1.3. The project’s other contributions embrace developing METHODOLOGY FOR CROSS-LINGUISTIC DATA COLLECTION AND ANALYSIS, including cross-linguistic word space analysis without parallel data, and a meta-language for semantic comparison.

2. *Preliminaries*

Temperature perception in humans involves two types of universal experience: temperature sensation/evaluation of the temperature of other entities, based on perception received by the skin, and thermal comfort, each with its own scale and reference point. Temperature phenomena are relatively easily perceptible by humans and are crucial for them, but their conceptualisation involves a complex interplay between external reality, bodily experience and evaluation of the relevant properties with regard to their functions in the human life. The meanings of temperature terms are, thus, both embodied and perspectival. Rather than reflecting the external world objectively, they offer a naïve picture of it, permeated with folk theories that are based on people’s experience and rooted in their culture (cultural models).

In talking about temperature perception, we can distinguish three or even four subdomains, in which languages vary as to how much these distinctions are reflected in the morphosyntax/lexical choice (cf. also Plank 2003). In Table 1 for **predicative constructions**, English uses the same adjective *cold* in the same construction in all the cases, but has additionally a verb, *freeze*, for experiencer-based temperature. German goes further in also having a special syntactic construction here. French has separate constructions for ambient and experiencer-based temperatures, while Japanese combines the choice between two lexemes for ‘cold’ with the choice between two syntactic patterns. Temperature terms can belong to different word classes presumably due to the fact that temperature varies “as to its time-stability, depending on who or what it is attributed to” (Plank 2003). Languages differ as to how many temperature terms they have and how these categorize, or carve up each of the different temperature subdomains or the temperature domain in general. The main query here is the extent to which these cross-linguistically different categorizations are still commensurable with each other, i.e., the allowed / attested “proportion” between universal and language-specific, or arbitrary categorization.

Table 1: Talking about ‘cold’ in English, German, French and Japanese.

	Experience-based temperature			Experiencer-based temperature
	Tactile-temperature	Non-tactile-temperature		
		quasi-referential	ambient	
English	<i>The stones are cold.</i>	<i>The wind is cold.</i>	<i>It is cold today.</i>	<i>I am cold.</i> <i>I am freezing.</i>
German	<i>Die Steine sind kalt.</i>	<i>Der Wind ist kalt.</i>	<i>Es ist kalt heute.</i>	<i>Mir ist kalt.</i> <i>Ich friere.</i>
French	<i>Les pierres sont froides.</i>	<i>Le vent est froid.</i>	<i>Il fait froid.</i>	<i>J’ai froid.</i>
Japanese	<i>Ishi ga (SUB) tsumetai.</i>	<i>Kaze ga (SUB) tsumetai.</i> <i>Kaze ga (SUB) (?) samui.</i>	<i>Kyô wa (TOP) samui.</i>	<i>(Watashi wa (TOP)) samui</i>

Temperature meanings are often semantically related to other meanings, either synchronically (within a polysemantic lexeme) or diachronically. Thus, as concepts rooted in basic and easily perceptible experiences, they often serve as source domains for various metaphors (‘warm feelings’, ‘hot news’) and are extended to other perceptual modalities (‘hot spices’, ‘warm colour’). Conversely, temperature meanings can develop from others, e.g., “prototypical” entities or activities with certain temperature characteristics (‘burn, fire’ > ‘hot’). Finally, the meanings of temperature terms can also change within the temperature domain itself, e.g. ‘warm, hot’ > ‘lukewarm’, as in Lat. *tep-* ‘warm’, Sanskrit *tâpas* ‘heat’ vs. English *tepid* ‘lukewarm’, or the interesting cases of diachronic enantiosemey (opposite meanings) like Swed. *sval* ‘cool’ vs. German *schwül* ‘stiffy, unpleasantly warm’.

3. Survey of the field

The integrated lexical-typological research on the temperature domain envisaged in the current proposal has roots in several different research traditions.

3.1. Lexical typology

Cross-linguistic research on categorization within conceptual domains is based on the idea that human experience is not delivered in pre-packed units and categories, but is chunked and categorized by humans themselves. Experiences that systematically correspond to the same word are taken to be perceived as instances of one category. The key concern in all this research is whether semantic categories are completely free to vary across languages, i.e. whether the human mind can carve the world at an infinite and arbitrary number of places, or whether there are limits to this, given by the clustering of ontological features of the real world, the organization of our

neurophysiology or of the human mind, or anything else. Only a few conceptual domains typically encoded by words have been subject to systematic cross-linguistic research on their semantic categorization, primarily COLOUR, BODY, KINSHIP, PERCEPTION, MOTION, EVENTS OF BREAKING AND CUTTING, DIMENSION and POSTURE.

Within cross-linguistic research on lexicon-grammar interaction, two traditions are particularly important for the proposal: work on word classes with close attention to lexical semantics and against the background of relatively fine-grained lexical distinctions (e.g. Dixon and Aikhenvald 2004), and the grammatical properties of words from a particular conceptual domain (e.g. Dahl & Koptjevskaja-Tamm 2001).

Systematic cross-linguistic research on semantic derivation and motivation (patterns of polysemy and semantic change) has so far been very limited. The best studied cases include BODY and the associations between PERCEPTION and COGNITION (e.g., Sweetser 1990, Evans & Wilkins 2000). Both demonstrate a complex interplay between universal tendencies and language- or culture-specific factors. The relative meager attention attracted by these issues in cross-linguistic research is surprising given their central role in cognitive semantics. In spite of its impressive progress, this theoretical framework has operated with a limited number of languages and has modest empirically founded insights with regard to cross-linguistic comparison.

3.2. Linguistic research on temperature

3.2.1. Earlier work by other researchers

Temperature terms have, on the whole, received relatively little attention. Apart from a few studies of temperature terms in particular languages or in two languages as contrasted to each other, there are very little empirical work and generalizations here, especially, on how the whole temperature domain is carved up by the temperature terms in a language (cf. Sutrop 1998 on Estonian, Shimotori 2004 on Japanese and Koptjevskaja-Tamm & Rakhilina 2006 on Swedish and Russian). Cross-linguistic research on temperature is mainly restricted to Sutrop (1998, 1999) and Plank (2003), which focus on how many basic temperature terms there are in a language (“Probably there are only 2-term, 3-term, or 4-term systems of basic terms”) and how they carve up the domain among themselves. Since the different criteria for identifying basic terms may not cluster, Sutrop and Plank occasionally suggest different treatments of one and the same phenomena and counter-intuitive classifications of languages.

In theoretical semantics, temperature adjectives have figured in discussions of lexical fields, antonymy and linguistic scales, where the main question is whether the temperature adjectives in a language are organized as points on one or two scales (cf. Lehrer 1970, Cruse & Togia 1995, Sutrop 1998). Within cognitive linguistics, temperature is touched upon by Clausner & Croft (1999) as an example of a locational domain. Most of these discussions depart from a single language, typically English, and give hardly any clues for understanding how and why the temperature-term systems vary across languages. Within the Natural-Semantic Metalanguage, Goddard (2001) discards temperature meanings as lexico-semantic universals, while Goddard & Wierzbicka (2006) suggest the general formula for describing the language-specific meanings of temperature terms via reference to fire (a part of evidence in both papers come from Koptjevskaja-Tamm & Rakhilina 1997, 2006).

There has been no systematic cross-linguistic research on the grammatical behaviour of temperature expressions. Dixon (2004) mentions ‘hot’ in the category of PHYSICAL PROPERTY, which might be lexicalized as adjectives in languages with medium-sized and large adjective classes. The list of expressions in Bossong’s (1998) study of the experienter encoding in the European languages includes “I am cold”.

Extended uses of temperature words have been studied indirectly in cognitive linguistics, primarily in research on the allegedly universal metaphors underlying emotions, e.g. AFFECTION IS WARMTH (Lakoff & Johnson 1997:50) and ANGER IS HEAT (Kövecses 1995, also Goossens 1998). An important question raised in Geeraerts & Grondelaers (1995) is to what degree such extensions reflect universal metaphorical patterns or are based on common cultural traditions. In any case, the current empirical evidence for the suggested metaphors is very meagre. Several studies on particular languages focus on or just mention extended uses of temperature words in reference to emotions, to other perception modalities (taste, colour, sound), richness/poverty, or quickness/slowness (e.g., Abelin 1988 and Bergström 2002 for Swedish, Shimotori 2004 and Shindo 1998, 1999 for Japanese). Sutrop (1999) and Plank (2003) offer a few remarks on the origin of temperature terms and their semantic change.

3.2. Own research and preliminary results

MKT's own earlier cross-linguistic research on temperature terms is documented in Koptjevskaja-Tamm & Rakhilina (1999, 2006) and less directly in Koptjevskaja-Tamm (forthc.) and Koptjevskaja-Tamm et al. (2007). MKT has also supervised a MA thesis on Japanese (Shimotori 2004), collected preliminary data on seven languages and has had less systematic discussions of temperature terms with language experts. Some of the preliminary results and hypotheses are as follows:

1. Temperature qualification of nominals is partly determined by the physical properties of the objects they denote, but even more by their functions in human life and by the cultural models associated with them. Temperature concepts are much more sensitive to the meanings of the entities they describe than is normally assumed, which has theoretical and methodological implications for the issues of lexical relations (e.g. antonymy) and basic terms. It is e.g. methodologically safer to treat the different basic-term criteria separately rather than to lump them together.
2. The meaning of temperature terms is defined according to parameters that are salient for humans, are distinguishable by simple procedures relating to the human body and have only approximate physical correlates, e.g. tactile vs. non-tactile perception, unpleasantly hot vs. other temperatures, etc. While the addition of languages to the study most probably implies addition of new temperature parameters or modification of the suggested ones, it is hypothesized that this process will end up at some point and that the resulting lists of parameters will be relatively short.
3. Parameters can interact in various ways. E.g. while the distinction for tactile and non-tactile perception in Russian is salient for the 'hot' temperatures, a similar distinction in Japanese is most robust for the 'cold' temperatures (Shimotori 2004). An important issue is to what extent there are implications among the parameters.
4. Semantic distinctions reflected in lexical choice in some languages can correspond to morphosyntactic differences in other languages. Central issues here are the cross-linguistic "division of labour" between the two linguistic sub-systems and the question of "semiotic ecology", i.e. how semantic choices made in one subsystem affect those in others (e.g., will there be any differences in the semantic elaboration of the temperature domains with temperature terms lexicalized as adjectives vs. verbs?). This is one of the reasons for a combined research on lexicalization of temperature concepts and on lexicon-grammar interaction within this domain.
5. Extensions from the temperature domain are dependent on its categorization. E.g., the border between *varm* 'warm' and *het* 'hot' in Swedish is higher than the one between *teplyj* and *gorjachij* in Russian, which is kept in their extended uses in the emotional domain: *het* is used for passions, while *gorjachij* has a wider applicability.

It is therefore reasonable to combine research on semantic derivation and motivation relevant for the temperature domain with a detailed study of its categorization.

6. Closely related languages can show remarkable differences in their uses of temperature adjectives, even when these are cognates to each other (e.g., Russian is the only Slavic language to have a consistent distinction between tactile and non-tactile 'hot'); conversely, temperature systems can show remarkable areal patterns.

7. Lexicalization of temperature terms as adjectives seems to be relatively rare.

8. While some languages show extensive semantic derivation from the temperature domain, others lack it or use it to a limited degree (e.g., Pirahã in Brasil, the Oceanic and Australian Aboriginal languages). Languages vary as to which temperature term has predominantly positive associations in its extended use (cf. 'cold' in Wolof vs. 'warm' in the European languages), partly due to the different climatic conditions.

4. Project description

As stated in Section 1, the project is envisaged as an integrated lexical-typological study of the temperature domain across a number of genetically, areally and structurally diverse languages with the aim of describing and accounting for the cross-linguistic variation within the domain from three different angles – lexicalization of temperature concepts, lexicon-grammar interaction, and semantic derivation / motivation. The integrated approach advocated here has hardly ever been practiced earlier, with the possible exception of the project “The human body and space” at Max-Planck Institute for Psycholinguistics (Nijmegen) and the project “Motion in liquid medium” at Moscow to which MKT has contributed. The project is anchored in functionally and cognitively oriented linguistic theories, but is not limited to a particular one. Of primary relevance for data analysis, representation of the results and explanations are Basic Linguistic Theory (the standard in modern grammatical typology), non-formal semantic theories – Cognitive Linguistics (most importantly, Frame Semantics and the Theory of Conceptual Metaphors), Natural Semantic Metalanguage, Moscow School of Semantics – and Cultural Linguistics.

The primary methodological concerns for the project are a) methods of data collection, b) cross-linguistic identification of studied phenomena and their (semantic) analysis, and c) the meta-language used for representing meanings.

a) Secondary data sources, such as dictionaries, grammars and studies of the relevant phenomena in particular languages, are of limited value in lexical typology since the data is normally incomplete and scattered across different sources, dictionaries provide vague and circular definitions, etc. Data collection and analysis in the project will mainly rest on elicitation of primary data and corpus search, occasionally complemented by secondary sources. Combining the different methods the project aims at covering the temperature domains of 25-30 languages.

Elicitation of primary data for the envisaged project will follow the guidelines that have already been discussed at several workshops, tested on several languages and revised on the basis of the input from the researchers testing them. The guidelines include naming and ranking tasks, checklists based on the ideas of canonical cases/uses (partly inspired by the Natural Semantic Metalanguage and the FrameNet) with detailed semantic and grammatical instructions, and a number of open questions, www.ling.su/staff/tamm/tempquest.pdf. The guidelines are already used as the basis for systematic and coherent descriptions of the temperature domain meant for a collaborative volume by the following internationally known experts: Cliff Goddard (Malay), Farzad Sharifian (Persian), Ning Yu (Mandarin), Prashant Pardeshi (Marathi), Michael Daniel (Archi), Päivi Juvonen (Finnish), Martine Vanhove and

Guillaume Ségerer (French), Felix Ameka (Ewe), Olga Khanina (Enets), Valentin Goussev (Nenets), Bernard Carron (Hausa), Loïc-Michel Perrin (Wolof), Samuel Atintono (Gur languages), Eva Rothmaler (Kanuri), Frank Seifart (Amazonian languages). In addition, the guidelines will be used for a broader data collection by MKT and the project assistant (the preliminary list includes Slavic, Romance and Germanic, Hungarian, Saami, Samoan, Amharic, Arabic, Hebrew and Turkish).

Corpus search will primarily involve word-space modeling, which is particularly suitable for the proposed project, because it allows semantic similarities between words to be automatically extracted from text corpora, requiring a minimum of human effort and linguistic preprocessing (Sahlgren 2006). By accumulating word spaces from text data, we can study how temperature terms cluster and form semantic networks in far more languages than would be possible to investigate manually, which is a considerable advantage in typological linguistic investigations. Since the only requirement is that there are sufficiently large corpora in the languages under investigation, our corpus study will focus on major languages: there are ample corpora for most European languages, for major Asian languages (e.g. Japanese and Mandarin), for some Middle-Eastern (e.g. Arabic) and some African (e.g. Amharic) languages. Since such an extensive application of word spaces has never been attempted before, the corpus study will contribute not only to typological research, but also to word-space research itself, and to our understanding of the applicability of distributional models to typologically diverse languages.

b) Although much of cross-linguistic comparison is based on meaning seen as denotation, the data collection techniques used in the project aim at discovering descriptive meanings. Decisions on what count as several meanings of a lexeme vs. one more general meaning normally require elaborated tests, difficult even within one language and often leading to a huge proliferation of the meanings within a lexeme. This richness is rarely of interest for typological comparison, which tries to maintain a balance between language-specific details and cross-linguistic generalizations. Our data collection techniques provide a number of contexts, or “an etic grid” for capturing possible distinctions within a domain, with the results that the meaning of a word can easily become reduced to the set of its uses. The logical step from an ETIC definition to finding out the commonalities behind the different uses and, ideally, arriving at a reasonable characterization of the descriptive meaning will go hand in hand with deciding what constitutes one meaning. However, in accordance with the current practices of cross-linguistic lexical studies we will take a pragmatic stance on this issue, choosing the solution that suits best the questions asked in the study.

c) A central complication for cross-linguistic studies on the lexicon is the problem of a consistent meta-language for representing meanings within and across languages, partly due to the general enormous gap between theoretical semantics and theoretical lexicology, on the one hand, and actual lexicographic practices. We intend to test and refine meaning representations and explications inspired by the Constructional Grammar and the Natural Semantic Metalanguage.

Timetable

2009: a pilot study, testing and refining the methodology of data collection, working out and testing a cross-linguistically applicable model of semantic analysis. A workshop with the language experts using the guidelines.

2100 data collection for the whole sample

2110: analysis of results, preparing the final publication

Deliverables for the project will include a collective monography and a series of journal publications.