

## *Studies in Language Companion Series (SLCS)*

This series has been established as a companion series to the periodical *Studies in Language*.

### **Editors**

Werner Abraham  
University of Vienna

Michael Noonan  
University of Wisconsin, Milwaukee

### **Editorial Board**

Joan Bybee  
University of New Mexico

Christian Lehmann  
University of Erfurt

Ulrike Claudi  
University of Cologne

Robert E. Longacre  
University of Texas, Arlington

Bernard Comrie  
Max Planck Institute, Leipzig  
University of California, Santa Barbara

Brian MacWhinney  
Carnegie-Mellon University

William Croft  
University of New Mexico

Marianne Mithun  
University of California, Santa Barbara

Östen Dahl  
University of Stockholm

Edith Moravcsik  
University of Wisconsin, Milwaukee

Gerrit J. Dimmendaal  
University of Cologne

Masayoshi Shibatani  
Rice University and Kobe University

Ekkehard König  
Free University of Berlin

Russell S. Tomlin  
University of Oregon

### **Volume 106**

From Polysemy to Semantic Change. Towards a typology  
of lexical semantic associations  
Edited by Martine Vanhove

# **From Polysemy to Semantic Change**

Towards a typology of lexical semantic associations

*Edited by*

**Martine Vanhove**

Llacan (Inalco, CNRS), Fédération TUL

John Benjamins Publishing Company

Amsterdam / Philadelphia

**2nd proofs**



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

#### Library of Congress Cataloging-in-Publication Data

From polysemy to semantic change : towards a typology of lexical semantic associations / edited by Martine Vanhove.

p. cm. (Studies in Language Companion Series, ISSN 0165-7763 ; v. 106)

Includes bibliographical references and index.

1. Semantics, Historical. 2. Polysemy. 3. Typology (Linguistics) I. Vanhove, Martine.

P325.5.H57F76 2008

401'.43--dc22

2008031821

ISBN 978 90 272 0573 5 (Hb; alk. paper)

© 2008 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands  
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

## Table of contents

Semantic associations: A foreword <i>Martine Vanhove</i>	vii
<b>Part 1. State of the art</b>	
Approaching lexical typology <i>Maria Koptjevskaja-Tamm</i>	3
<b>Part 2. Theoretical and methodological issues</b>	
Words and their meanings: Principles of variation and stabilization <i>Stéphane Robert</i>	55
The typology of semantic affinities <i>Bernard Pottier</i>	93
Cognitive onomasiology and lexical change: Around the eye <i>Peter Koch</i>	107
Mapping semantic spaces: A constructionist account of the “light verb” <i>xordaen</i> “eat” in Persian <i>Neiloufar Family</i>	139
Semantic maps and the typology of colexification: Intertwining polysemous networks across languages <i>Alexandre François</i>	163
A catalogue of semantic shifts: Towards a typology of semantic derivation <i>Anna A. Zalizniak</i>	217
Semantic associations and confluences in paradigmatic networks <i>Bruno Gaume, Karine Duvignau &amp; Martine Vanhove</i>	233
<b>Part 3. Case studies</b>	
About “Eating” in a few Niger-Congo languages <i>Emilio Bonvini</i>	267

# Approaching lexical typology

Maria Koptjevskaja-Tamm  
University of Stockholm

The paper aims at situating the research direction presented in the volume within the larger domain of typological research in general. It gives a short summary of what is meant by typological research, discusses the relation between semantic and lexical typology and the general premises for lexical-typological research. The bulk of the paper is devoted to the three main lexical-typological research foci – what meanings can(not) be expressed by a single word, what different meanings can be expressed by one and the same lexeme or by words derivationally related to each other, and what cross-linguistic patterns there are in lexicon-grammar interaction. The paper ends up with the general discussion of the urgent methodological problems facing lexical typology as a field.

**Keywords:** lexical typology; lexicon-grammar interaction; linguistic categorization; motivation; semantic typology

## 1. Introduction

The aim of the present article is to situate the research direction presented in the volume within the larger domain of modern typological research in general. As witnessed by the title of the volume, “From polysemy to semantic change: Towards a typology of lexical semantic associations”, three key words are crucial here – *typology*, *lexical* and *semantic*. The paper starts by a short summary of what is meant by typological research in general, then discusses the relation between semantic and lexical typology and points out several different groups of questions asked within lexical-typological research, its different foci. Section 3 touches on the general premises for lexical-typological research – possible words, semantic generality vs. polysemy, and the meaning of “meaning”. Sections 4–6 are devoted to the three main lexical-typological research foci introduced in Section 2 – what meanings can(not) be expressed by a single word, what different meanings can be expressed by one and the same lexeme or by words derivationally related to each other, and what cross-linguistic patterns there are in lexicon-grammar interaction. Each of these chapters considers numerous examples of research and the various methodological and theoretical issues relevant for it. The whole paper ends up with the general discussion of the urgent methodological problems facing lexical typology as a field.

## 2. Typology, semantic and lexical typology

The term “typology”, as is well known, has many different uses. What primarily matters for the present volume is typology understood as “the study of linguistic patterns that are found cross-linguistically, in particular, patterns that can be discovered solely by cross-linguistic comparison” (Croft 1990: 1). Typology can also refer to typological classification of languages into (structural) types on the basis of particular patterns for particular phenomena. Typological research is driven by the persuasion that the variation across attested (and, further, possible) human languages is severely restricted, and aims therefore at unveiling systematicity behind the whole huge complex of linguistic diversity. In pursuing their tasks, typologists raise – and often try to answer – important theoretical questions, such as the ones listed below.

- According to what parameters does a specific phenomenon vary across languages, in what patterns do these parameters (co-)occur?
- What generalisations can be made about attested vs. possible patterns?
- What is universal vs. language particular in a given phenomenon, what phenomena are frequent vs. rare?
- How are various linguistic phenomena distributed across the languages of the world?
- Which phenomena are genetically stable and which are subject to contact-induced change?
- How can the attested distribution of the different patterns across languages be explained?
- How can the attested cross-linguistic patterns/generalizations be explained?

The papers in the present volume do in fact focus on linguistic patterns that can be discovered only by cross-linguistic comparison – cross-linguistically recurrent patterns of polysemy, heterosemy and semantic change – and are therefore examples of typological research. The domain of research shared by the papers in the volume is, however, somewhat outside of the main interests of modern typological research, that has so far primarily focused on grammatical and, to a lesser degree, phonetic/phonological phenomena under the labels of “grammatical typology”, “syntactic typology”, “morphological typology”, “morphosyntactic typology” (or, quite often, just “typology”), “phonetic typology” and “phonological typology”. None of those would suit the direction of the volume. We are dealing here, first, with lexical and, second, with semantic phenomena – which are the primary objects of lexical vs. semantic typology.

The terms “semantic typology” and “lexical typology” are often used as if there were self-explanatory, but are only rarely explicitly defined. According to Evans (forthc.), semantic typology is “the systematic cross-linguistic study of how languages express meaning by way of signs”. What can be meant by lexical typology is, however,

less clear, apart from the evident fact that it involves cross-linguistic research on the lexicon. Many linguists will probably agree with Lehrer’s (1992: 249) widely quoted definition that lexical typology is concerned with the “characteristic ways in which language [...] packages semantic material into words” (cf. the excellent overviews in Koch 2001 and Brown 2001). Viewed as such, lexical typology can be considered a sub-branch of semantic typology concerned with the lexicon, as this is done in Evans (forthc.). Other definitions of lexical typology focus on “typologically relevant features in the grammatical structure of the lexicon” (Lehmann 1990: 163) or on typologically relevant vs. language-specific patterns of lexicon-grammar interaction (Behrens & Sasse 1997).

I think that a reasonable way of defining what can be meant by “lexical typology” is to view it as the cross-linguistic and typological dimension of lexicology. The probably most updated overview of lexicology as a field is found in the two volumes (Cruse et al. eds. 2002, 2005), the title of which (“Lexicology. An international handbook on the nature and structure of words and vocabularies”)

underlines the special orientation towards the two core areas which makes of lexicology an autonomous discipline, namely, the characterization of words and vocabularies, both as unitary wholes and as units displaying internal structure with respect both to form and content (Cruse et al. eds. 2002, 2005: viii–ix).

In the same vein as lexicology in general is not restricted to lexical semantics, lexical typology can include phenomena that are not of primary interest for semantic typology. Likewise, since lexicology is not completely opposed to either phonetics/phonology, morphology or syntax, cross-linguistic research on a number of word- and lexicon-related phenomena is – or can be – carried out either from different angles and with different foci, or within approaches that integrate several perspectives, goals and methods.

There are different kinds and groups of questions that can be addressed in typological research on words and vocabularies, or lexical typology, and that can therefore be considered as the different foci of lexical typology. Some of them (the most important ones, as I see them now) are listed below, but there are undoubtedly many others. The questions are often interrelated with each other, a point that will be stressed in the ensuing presentation.

- What is a possible word, or what can be meant by a word? Possible vs. impossible words in different languages, different criteria for identifying words and interaction among them, universal vs. language-specific restrictions on possible, impossible, better and worse words.
- What meanings can and cannot be expressed by a single word in different languages? Lexicalizations and lexicalization patterns, “universal” vs. language-specific lexicalizations, categorization within, or carving up of lexical fields/semantic

domains by lexical items, the architecture of the lexical fields/semantic domains (e.g., basic words vs. derived words).

- What different meanings can be expressed by one and the same lexeme, by lexemes within one and the same synchronic word family (words linked by derivational relations) or by lexemes historically derived from each other? Cross-linguistically recurrent patterns in the relations among the words and lexical items in the lexicon – a huge and heterogeneous category with many different subdivisions, a large part of which can be subsumed under the various aspects of motivation (for details see Koch 2001: 1156–1168 and Koch & Marzo 2007), e.g., semantic motivation (polysemy, semantic associations/semantic shifts) and morphological motivation (derivational patterns, including compounding).
- What cross-linguistic patterns are there in lexicon-grammar interaction?

The lexicon of a language is, of course, a dynamic and constantly changing complex structure where new words emerge, old words disappear or change in one or another way. Lexical-typological research has, thus, both *synchronic* and *diachronic* dimensions. *Historically oriented lexical typology* studies semantic change, grammaticalization and lexicalization processes (the latter understood as “a process by which new linguistic entities, be it simple or complex words or just new senses, become conventionalized on the level of the lexicon”, Blank 2001: 1603) as examples of diachronic processes showing cross-linguistically recurrent patterns.

The lexicons of most languages show different layers of origin with many words coming from “outside” – as direct loans, loan translations, etc. A particularly interesting aspect of historical lexical typology is the search for cross-linguistically recurrent patterns in contact-induced lexicalization and lexical change, e.g., differences in borrowability among the different parts of the lexicon and the corresponding processes in the integration of new words, or patterns of lexical acculturation (i.e., how lexica adjust to new objects and concepts).

Lexical-typological research can also be more *local*, e.g., restricted to a particular lexical field, a particular derivational process, a particular polysemy pattern, or more *general*, with the aim of uncovering patterns in the structuring of the lexicon that are supposed to have a bearing on many essential properties of the language. The latter includes various approaches to the issues of “basic” vs. non-basic vocabulary, or suggestions as to how characterize, compare and measure the lexical-typological profiles of different languages. In fact, some people prefer using the term “typological” (e.g., typological properties) for referring to what is considered as the more essential, central, or general properties of a language. In this understanding, a large portion of cross-linguistic research on words and vocabularies will not count as typological (Lehmann 1990) (this applies, among others, to what is called “local” lexical-typological research immediately above) – we will briefly discuss this position in Section 6.

In the ensuing sections I will give examples of lexical-typological research along these different lines, try to make generalizations on the state of art and, finally, point out some recurrent problems and directions for the future. Because of space restrictions, issues such as possible words or degree of borrowability will only be allotted a brief mention here.

### 3. General premises: Words and meanings

#### 3.1 Possible words

As any introductory textbook in linguistics will tell us, “word” can denote different things. The huge issue of what can be meant by word, or what is a possible word across languages will only be touched upon in this paper – and mainly from the semantic point of view. Traditional morphological typology, with its focus on how much and what kind of morphology is allowed in words across languages (cf. isolating vs. polysynthetic languages, etc.), represents one way of comparing possible words cross-linguistically. The issue is, however, much more complex and requires a truly integrating approach, where morphological (and further grammatical), phonetic, phonological and semantic criteria are all relevant, as well as psycholinguistic considerations (holistic storage and processing), and sociolinguistic and pragmatic factors (e.g., degree of conventionalization). The important contributions include Aikhenvald & Dixon (2002) and the ongoing project “Word domains” within the research programme “Autotyp” (directed by Balthasar Bickel and Johanna Nichols, [http://www.uni-leipzig.de/~autotyp/projects/wd\\_dom/wd\\_dom.html](http://www.uni-leipzig.de/~autotyp/projects/wd_dom/wd_dom.html)) which both focus on words as phonological and grammatical domains;<sup>1</sup> a useful discussion of compounds as words across languages is found in Wälchli (2005: 90–134).

A new and promising way of investigating the question of “possible words” in a human language is laid out in Greville Corbett’s (2007, *forthc.*) “canonical approach” to inflection, that evaluates the various formal ways in which the word forms of one and the same lexeme can be related to each other. Since definitions of linguistic phenomena normally evoke several different criteria, the basic idea behind Corbett’s “canonical approach” in typology is to take definitions to their logical endpoint, where the different criteria converge and together produce the best, clearest, indisputable “canonical” instance of the phenomenon. Real “canonical” instances are rare or even unattested, but they serve as a point from which the actual phenomena can be evaluated in the theoretical space of possibilities. In the particular case of inflection we can,

1. Last visited on April 24th 2007.

thus, evaluate agglutination vs. flection, various kinds of syncretism or suppletion, and, in general, various kinds of exceptions in inflectional morphology.

Now, since much of the discussion in the present paper will deal with the link between words and their meanings, of primary concern here will be *words as carriers of lexical meanings*. Several interrelated distinctions are important for understanding the kinds of questions usually asked here and the controversies surrounding possible answers to them. These include the distinction between semantic generality and polysemy (Section 3.2.) and the various understandings of “meaning” – denotation vs. sense, and approximate vs. precise (Section 3.3.).

### 3.2 Semantic generality vs. polysemy, or when are meanings lexicalized?

A classical issue in lexical semantics concerns the distinction between *semantic generality* and *polysemy*. Consider Table 1 for the English and Russian verbs designating motion in water (aquamotion).

Table 1. Aquamotion verbs in English and Russian

	Passive motion (‘float’)	Self-propelled motion of animate figure (‘swim’)	Motion of vessels and people aboard (‘sail’)
English	<i>float</i>	<i>swim</i>	<i>sail</i>
Russian	<i>plyt’/plavat’</i>		

English distinguishes among three different verbs – *float*, *swim* and *sail*. *Float* designates passive motion; *swim* – active, self-propelled motion of animate figure, and *sail* – active motion of vessels and people aboard. To simplify matters, we can say that we have three different meanings, or concepts here – “float”, “swim” and “sail” – each of which exists as the meaning of a particular lexeme in English, or is *lexicalized*. Russian has two aquamotion verbs *plavat’/plyt’*, where the main difference is directionality of the designated motion – *plavat’* is multidirectional, while *plyt’* is unidirectional. Since each of these two verbs corresponds roughly to the same situations as the three English verbs together, we will for the sake of simplicity concentrate on just one verb, *plyt’*. The question is now whether the three meanings ‘float’, ‘swim’ and ‘sail’ are lexicalized in Russian. There are at least three theoretical and methodological possibilities here, e.g., semantic generality, polysemy and agnosticism.

First, *semantic generality*: it could very well happen that *plyt’* is semantically general and does not distinguish among “float”, “swim” and “sail” at all. In that case we could say that Russian does not lexicalize the differences among “float”, “swim” and “sail” in having just one and the same word (or one word couple) covering all the three meanings.

The second possibility, *polysemy*, would mean that *plyt’* does in fact distinguish at least among the three different meanings “float”, “swim” and “sail”. In that case we

could still say that each of these meanings is lexicalized in Russian – however, not as the meaning of its “own” particular lexeme, but rather as the meaning of a particular *lexical unit*. A lexical unit is, in turn, defined as the pairing of a single specifiable meaning/sense with a lexical form (Cruse 1986: 77–78), so that a polysemous word is a lexeme consisting of several lexical units. Fig. 1 visualizes the difference between potential semantic generality vs. polysemy in the case of *plyt’*. Each of the double-headed arrows represents a lexical unit (LU); the lexical form {*plyt’*} covers all the inflectional forms that together constitute the formal side of this lexeme.

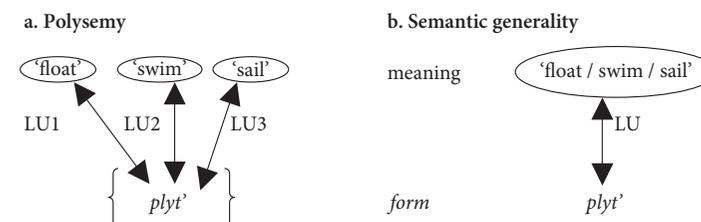


Figure 1. Representing polysemy (a) vs. semantic generality (b).

The third possibility is to leave aside the problem of semantic generality vs. polysemy and to remain *agnostic* about the correct semantic analysis of a particular word. This is the “default” interpretation of the data in Table 1. Under this view, what matters is the fact that Russian has only one lexeme (or, rather, a couple of directionality-related lexemes) corresponding to the three different English ones.

There are various tests for distinguishing between semantic generality and polysemy, e.g., the distinct meanings within a lexeme having different syntactic properties. On the basis of this particular test it can be argued that each of the two Russian verbs *plavat’/plyt’* distinguishes among several meanings, very much along the lines of the English system (cf. the analysis in Rakhilina 2006 and in Koptjevskaja-Tamm et al. forthc.), and that Fig. 1a gives a more faithful representation than Fig. 1b. Thus, when describing inanimate objects moving with water ( $\approx$  ‘float’), *plyt’* has to combine with an overt indication of the PATH normally expressed by a combination of *po* “on” and the reference to the surface, as in (1a); no indication of the SOURCE or GOAL of the motion is allowed. Whenever *plyt’* designates self-propelled motion of animate entities ( $\approx$  ‘swim’), it has to combine with an overt indication of either the SOURCE or GOAL of the motion, to the exclusion of the “surface” where it takes place, i.e., PATH (ex. 1b). Finally, when *plyt’* refers to the motion of vessels ( $\approx$  ‘sail’), it can take either indications of the SOURCE/GOAL of the motion, or of the surface (PATH); in addition, both can be combined in one and the same sentence, as in ex. (1c).

- (1) a. *Po rek-e ply-l-i želt-ye*  
 on river-DAT.SG “plyt”-PRET.PL yellow-NOM.PL  
*list'-ja (\*k bereg-u).*  
 leaf-NOM.PL (\*towards shore-DAT.SG)  
 ‘Yellow leaves were floating in the river (\*to the shore).’
- b. *On ply-l k bereg-u (\*po rek-e)*  
 he.NOM “plyt”-PRET.M.SG towards shore-DAT.SG (\*on river-DAT.SG)  
 ‘He was swimming towards the shore (\*in the river).’
- c. *Korabl' ply-l po Volg-e do Kazan-i*  
 ship.NOM.SG “plyt”-PRET.M.SG on Volga-DAT.SG to Kazan-GEN.SG  
 ‘The ship was sailing on the Volga river to the town of Kazan.’

Distinguishing between semantic generality and polysemy is, on the whole, a tricky business. Opinions on what polysemy amounts to and how to search for it differ considerably among different semantic theories and practices (such as dictionary entries), not to mention language users (see Riemer 2005 for a recent overview of the problem). In general, decisions on what should count as several meanings of one and the same lexeme vs. one more general meaning require sophisticated analyses and tests, difficult enough within one language, hard to carry out in several and impossible in many. The current practices of cross-linguistic lexical studies seem to take a fairly pragmatic stance on the choice between the three options represented by Table 1 and Fig. 1a-b, choosing the one that suits best the questions asked in a particular study (and also asking reasonable questions given the available data). Suppose that we are interested in the question of whether ‘swim’, ‘float’ and ‘sail’ belong to the *universally lexicalized meaning*. For this kind of question it hardly matters how exactly these meanings are expressed – as separate lexemes, as the different meanings of one and the same lexeme, or as words derivationally related to each other. So even if Russian does not have different words for ‘swim’, ‘float’ and ‘sail’, it still lexicalizes these meanings. On the other hand, the very fact that one and the same word *plyt'* corresponds to the three different words in English is very interesting too: according to the common-sense interpretation, driven by considerations of iconicity, this is an indication of their *semantic similarity*. Further interpretations of this first impression may differ – whether we are talking about the cross-linguistic variation in carving up, or categorizing the corresponding conceptual domain, on the one hand, or about lexical semantic associations, on the other – but the fact of the semantic affinity remains and has to be accounted for.

Consistent application of semantic tests may, in fact, lead to a huge proliferation and splitting of the meanings within a lexeme (cf. with some of the practices within the cognitive semantics). This richness is often of marginal interest for cross-linguistic and typological comparison, which is, by nature, reductionistic and tries to maintain a reasonable balance between language-specific details, on the one hand, and cross-linguistically applicable generalizations, abstractions and simplifications, on the

other hand. Here again, what is represented as one meaning vs. several meanings in cross-linguistic comparison tends to be governed by pragmatic considerations. To take an example, one of the meanings of the verb *to sail* (in combination with a human figure and a direct object) is “to control the movement of a boat or ship”, rather than just “to go in a boat or ship”. Even if this distinction is interesting per se, it can either be brought to the fore or neglected depending on what a particular study focuses on.

The method of semantic maps that has been successfully used in cross-linguistic comparison of grammatical semantics is explicitly agnostic about the distinction between polysemy and semantic generality (cf. Haspelmath 2003: 231), the same position is advocated by François (this volume) for cross-linguistic studies of lexical associations and is (implicitly) taken in cross-linguistic studies of categorization of conceptual domains based on multidimensional scaling (Levinson & Mejra 2003; Majid et al. 2007; Wälchli 2006/2007; see also Section 3.3. in the present paper).

### 3.3 The meaning of “meaning”: Denotation vs. sense, approximate vs. precise meaning definitions

The meaning of “meaning” is a key issue in semantics, where opinions vary. For our purposes the main and generally recognized partition goes between *denotation/extension* vs. *intension/(descriptive) meaning/sense*. The relations between the two are complicated, and the emphasis on the one or the other can have different implications for a cross-linguistic comparison.

Consider the domain of temperature. The Swedish adjective *ljummen* (“lukewarm”) covers a narrow and well-defined range of temperatures, “neutral” temperatures, those corresponding to the temperature of the human skin and feeling neither warm nor cold. The Russian adjective *teplyj*, whose standard translation into English is *warm*, also denotes a relatively restricted range of temperatures, the greater portion of which is covered by *ljum(men)*. From the denotational point of view, the two adjectives are, thus, fairly similar to each other, with the denotational range of *teplyj* slightly exceeding that of *ljummen*, as is schematically visualized in Fig. 2.

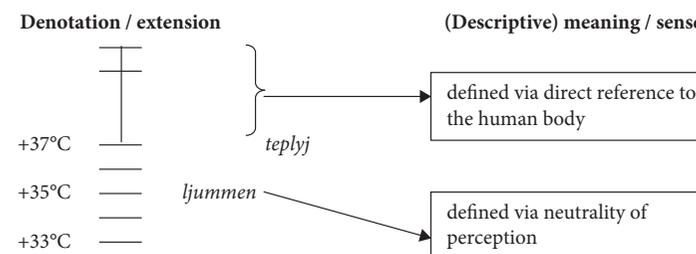


Figure 2. Extension/denotation vs. (descriptive) meaning/sense for the temperature adjectives *ljummen* (Swedish) and *teplyj* (Russian).

The descriptive meanings of the two adjectives, however, appear very different, as argued in (Koptjevskaja-Tamm & Rakhilina 2006: 259) and as also visualized in Fig. 2. First, *teplyj* does have a clear “warming” orientation: an entity qualified by its comparative form (*bolee teplyj, teplee*) has a HIGHER temperature than the one it is compared to. *Ljummen* often lacks this clear orientation. Thus, depending on pragmatic factors, an entity qualified by its comparative form (*mera ljummen, ljummare*) can have either a HIGHER (as in *Hans öl är ljummare än min* “His beer is more lukewarm than mine”) or a LOWER (as in *Mitt te har svalnat, hans är ännu ljummare* “My tea has cooled down, his is even more lukewarm”) temperature than the one it is compared to. In addition, *teplyj* is often used with body-part terms and has very positive connotations in metaphorical uses – e.g., *teplye slova, čuvstva, otnošeniya* “warm (positive, friendly) words, feelings, relations”. On the basis of these observations we have chosen to define the meaning of *teplyj* via direct reference to the (human) body. *Teplyj* designates temperatures that correspond to or are not significantly higher than the temperature of the human body/skin or that maintain the temperature of the human body without too much effort on the part of the human being, and therefore cause an agreeable sensation of comfort and coziness. *Ljummen*, on the other hand, never qualifies body-part terms and lacks any positive connotations in metaphorical uses – e.g., *ljumma kânslor, reaktioner* “weak, neutral feelings, reactions”. Thus, while *teplyj* covers temperatures that are “normal” with regard to the human body, that feel warm, but not exceedingly so, the meaning of *ljummen* lacks direct reference to the body, but evokes “neutrality” of perception.

Although for many serious semanticists, lexicographers and lexicologists semantic analysis stands for coming to grips with descriptive meanings, or senses, the enterprise is far from obvious even in the researcher’s native tongue and gets easily insurmountable in other languages. As a consequence, much of cross-linguistic comparison is based on meanings defined as denotations, with various methods for eliciting, defining and evaluating expression-denotation couples (pictures, videoclips, Munsell colour chips, etc.). In other words, the question of “what meanings can be or cannot be expressed by single words in a language” often amounts to “what are possible/impossible denotational ranges of single words in a language”. There are various reasons for why such approaches are insufficient, including Quine’s (1960: 29) famous “gavagai”-problem (if a person whose language you don’t know says *gavagai* when a white rabbit appears in front of you both, how can you be sure about what (s)he really means?). Another big problem is that many meanings – or, rather, many conceptual domains – hardly lean themselves to being investigated via denotation-based techniques: for instance, how do you get at the meaning of “think” or “love”? (cf. also Evans & Sasse 2007).

Other ways of dealing with the meaning in cross-linguistic lexical studies make use of “translational equivalents” found in dictionaries and word lists, questionnaires

and parallel corpora. These are, again, problematic in various ways. In particular, dictionaries are a favourite object of ridicule in theoretical work on semantics and lexicography, for providing vague and circular definitions. We will come back to the issue of approximate vs. precise meaning definitions (and cross-linguistic identity of meanings, cf. Goddard 2001: 2–3), which is, in fact, crucial in cross-linguistic studies.

Denotation-based techniques for data collection, questionnaires and parallel corpora effectively neglect the issue of semantic generality/polysemy, discussed in the previous section. They provide often a number of contexts, or “an etic grid” for capturing (logically) possible distinctions within a domain, with the results that the meaning of a word can easily become reduced to the set of its uses (an “etic definition”). The logical step from an *etic* definition to an *emic* one (i.e., finding out the commonalities behind the different uses and, ideally, arriving at a reasonable characterization of the descriptive meaning) goes hand in hand with deciding what constitutes one meaning, i.e., distinguishing between semantic generality and polysemy (cf. Evans *forthc.* for the discussion of etic vs. emic definitions in semantic typology).

A central complication for cross-linguistic studies on the lexicon – and, further, in most cross-linguistic research where meaning is involved – is created by the problem of a consistent meta-language for representing meanings within and across languages. This, in turn, is related to the general enormous gap between theoretical semantics and theoretical lexicology, on the one hand, and actual lexicographic practices. The most serious candidate on the market is the Natural Semantic Metalanguage, originally advocated by Anna Wierzbicka. The proponents of the NSM take polysemy very seriously, strive for comparing descriptive meanings rather than denotational ranges, and aim at providing precise meaning definitions by means of reductive paraphrases based on a principled set of “universal semantic primitives” (see, for instance, Goddard & Wierzbicka (eds.) 1994; Goddard 2001; Wierzbicka 1990, 2007). The theory has both positive and negative sides (cf., e.g., the discussion in Krifka (ed.) 2003; Riemer 2002 and Evans *forthc.*), but on the whole it enjoys less popularity and attention in the typological enterprise than it deserves.

#### 4. What meanings can and cannot be expressed by a single word?

What meanings can be or cannot be expressed by a single word in different languages, or what word meanings are universal, frequent, possible, impossible? Are there any universal (or at least statistically predominant) restrictions on the meanings that can or cannot be expressed by single words across languages, or are languages more or less free to choose here? This section will be devoted to research that focuses on the issues of cross-linguistically recurrent and typologically relevant aspects of lexicalizations and lexicalization patterns. Underlying it is the tension between two opposing

hypotheses on linguistic categorization, among others, on the concepts expressed by words. One hypothesis holds that categorization is universal, at least when it comes to basic, universal and daily situations, so that lexical meanings “originate in nonlinguistic cognition, and are shaped by perceptual and cognitive predispositions, environmental and biological constraints, and activities common to people everywhere” (Majid et al. 2007: 134). The other suggests that lexical meanings “do not reflect shared nonlinguistic cognition directly, but are to some extent linguistic conventions that are free to vary – no doubt within limits – according to historical, cultural, and environmental circumstances” (Majid et al. 2007: 134, see also the references there).

In this section we will look at studies dealing with categorization within, or carving up of lexical fields/semantic domains by lexical items.

#### 4.1 Categorization within lexical fields and conceptual domains:

##### A couple of examples to start with

The basic idea underlying cross-linguistic research on categorization within lexical fields and conceptual domains (coherent segments of experience and knowledge about them) is that human experience is not delivered in nicely pre-packed units, categories and types, but has to be chunked, organized and categorized by human beings themselves. Categories correspond to experiences that are perceived to have features in common. When experiences are systematically encoded by one and the same linguistic label (e.g., by the same word) they are, most probably, perceived as being fairly similar to each other; that is they are taken to represent one and the same class, or to correspond to one and same concept or lexical meaning.

A simple example of what can be meant by different ways of categorizing, or carving up a conceptual domain across languages is given in Table 2, which shows how the inventories of body-part terms in six languages differ in the extent to which they distinguish between hand vs. arm, foot vs. leg, and finger vs. toe by conventionalised, lexicalized expressions (“labels”).

Table 2. Hand vs. arm, foot vs. leg, finger vs. toe in English, Italian, Rumanian, Estonian, Japanese and Russian

English	Italian	Rumanian	Estonian	Japanese	Russian
hand	<i>mano</i>	<i>mină</i>	<i>käsi</i>	<i>te</i>	<i>ruka</i>
arm	<i>braccio</i>	<i>brat,</i>	<i>käsi(vars)</i>	<i>ude</i>	
foot	<i>piede</i>	<i>picior</i>	<i>jalg</i>	<i>ashi</i>	<i>noga</i>
leg	<i>gamba</i>				
finger	<i>dito</i>	<i>deget</i>	<i>sõrm</i>	<i>yubi</i>	<i>palec</i>
toe			<i>varvas</i>		

The table above follows the same practice of representing “lexicalization” in a fairly unsophisticated way as Table 1 in Section 3.2., without asking the question of whether *ruka* in Russian or *yubi* in Japanese are polysemous or semantically general. What matters here is simply how many different lexemes there are and how they partition the domain. A somewhat more complicated example, partly introduced in Section 3.2., is given in Table 3, which shows the verbs used for talking about water-related motion (aquamotion) in three languages – Swedish, Dutch and Russian. The table includes both motion of water itself (“flow” in English) and motion/location of other entities (other figures) with water as ground. Here, again, the Russian verbs *plyt’/plavat’* are treated as one semantic unit, rather than two sets of different senses. *Flyta* in Swedish appears, however, at two different places – this does not per se imply any strong conviction that the case is much different from the Russian verb couple, but demonstrates problems with two-dimensional representations.

Table 3. A part of the aquamotion domain in Swedish, Dutch and Russian (based on Koptjevskaja-Tamm et al. forthc.)

	Active motion of an animate Figure	Motion of vessels and people aboard				Passive motion; location on water		Motion of water
		Sailing boats	Motor-driven vessels	Rowing boats	Canoes	Motion out of control	Neutral motion/ Location	
Sw	<i>simma</i>	<i>segla</i>	(no specific aquamotion verbs)	<i>ro</i>	<i>paddla</i>	<i>driva</i>	<i>flyta</i>	<i>flyta, rinna</i>
Du	<i>zwemmen</i>	<i>zeilen</i>	<i>varen</i>	<i>roeien</i>	<i>paddelen</i>	<i>drijven</i>		<i>stromen</i>
Ru		<i>plyt’/plavat’</i>		<i>gresti</i>		<i>nestis’</i>		<i>teč, lit’sja</i>
		<i>pod parusami</i>						

As these examples show, languages differ considerably as to how many different lexemes they have for talking about comparable domains and how exactly these words partition the domains. It is therefore reasonable to ask whether there is any systematicity underlying the obvious cross-linguistic variation. Whatever the answer is, it requires explanation.

Only a handful of conceptual domains typically encoded by words (rather than by grammatical means) 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 not considered here).<sup>2</sup> The list can be made slightly longer, if we include words and expressions with more grammatical meanings, such as INDEFINITE PRONOUNS (Haspelmath 1997), various QUANTIFIERS (cf. Bach et al. 1995; Auwera 2001; Gil 2001), INTERROGATIVES (Cysouw 2004), PHRASAL ADVERBIALS (Auwera 1998) and SPATIAL ADPOSITIONS (Levinson & Meira 2003) – these won't be considered below.

#### 4.2 Domain-categorization studies: Language coverage and focus

The standard textbook example of underlying systematicity behind the striking cross-linguistic diversity, COLOUR, remains, probably, the most widely researched on domain in lexical typology, in terms of the languages covered in systematic comparison by means of comparable and elaborated methodology, and the intensity, diversity and depth of theoretical discussions. Kay & Maffi's (2005) chapter on COLOUR in the *World Atlas of Language Structures* is based on 119 languages, the data coming primarily from Berlin, Kay & Merrifield's World Color Survey in 1976–78, but there are many more languages the colour inventories of which have been subject to systematic research and which have figured in linguistic discussions (the relevant literature is too extensive for being listed here, cf. MacLaury 1997, 2001 and Payne 2006 and the references there, also <http://www.icsi.berkeley.edu/wcs/for> the *World Color Survey Site*).

KINSHIP terminologies have for a long time been a favourite semantic field among anthropologists and anthropologically oriented linguists. Detailed and systematic descriptions of the domain are available for many hundreds of languages, and there is a long tradition of classifying the resulting systems into a small number of types. As a rule, such classifications do not consider the whole kinship systems, but concentrate only on their subparts. Probably the cross-linguistically most systematic study, Nerlove & Romney (1967), focuses on sibling terminologies in 245 languages.

For the BODY domain, the most comprehensive – in terms of the number and representativity of the included languages – studies are the two chapters by Brown (2005a-b) in the *World Atlas of Language Structures*. One of them classifies 617 languages according to whether they use the same or different words for hand and arm, the other one asks the same question for finger and hand in 593 languages. Cross-linguistic studies on categorization of the whole body cover a handful of languages – the milestones here are Brown (1976), Andersen (1978) and, in particular, Majid et al. (eds.) (2006), with detailed and systematic studies of ten languages.

The modern lexical-typological research on MOTION verbs is for many people firmly associated with the tradition stemming from Talmy's seminal chapter (1985),

2. For the latest developments on the latter concept see Ameka & Levinson 2007, cf. also Newman 2002.

which focuses on the components of a motion event that are systematically encoded within motion verbs in different languages (lexicalization patterns). It is not quite clear how many languages have been systematically studied from this point of view – my impression is that they are not so many. Wälchli's on-going research on motion events in general covers more than 100 languages (Wälchli 2006; Wälchli & Zúñiga 2006). Since the motion domain is, in fact, very complex and heterogeneous and is normally encoded by many different linguistic means, it is reasonable to split it into smaller sub-domains for meaningful and doable cross-linguistic studies. Ricca (1993) focuses on the distinction between the DEICTIC MOTION verb ("come") and the NON-DEICTIC verb ("go") in twenty European languages, whereas the papers in Maisak & Rakhilina (eds.) (2007) present detailed and systematic studies of AQUAMOTION verbs across 40 genetically, structurally and areally diverse languages.

The most systematic cross-linguistic study of PERCEPTION verbs has been carried out by Viberg (1984, 2001) on the basis of fifty languages (and further confirmed by the Australian languages in Evans & Wilkins 2000), whereas the domain of CUTTING AND BREAKING EVENTS (Majid & Bowerman eds. 2007) has been investigated in 28 genetically and structurally diverse languages (covering 13 languages families, four isolates and a creole language). Finally, Lang (2001) presents a cross-linguistic comparison of DIMENSION terms (like "wide", "long", etc.) in a sample of forty languages coming from Europe and Asia.

#### 4.3 Methodology

Research on domain-categorization – as lexical typology in general – has on the whole made relatively little use of secondary sources, but often relies on primary data. Some of the studies on KINSHIP terminology (e.g., Nerlove & Romney's 1967; or Greenberg 1980) largely utilize the available earlier descriptions of particular languages, but this, in turn, depends on the long descriptive tradition of the domain starting with Morgan's (1870) early typology of kinship systems. Viberg's study (1984, 2001) of PERCEPTION verbs involves a combination of secondary data sources (dictionaries, word lists and more detailed descriptions) and systematic translation of sentences in a questionnaire; Ricca's (1993) study of DEICTIC verbs is mainly based on a questionnaire.

The majority of cross-linguistic COLOUR studies have all involved elaborated and systematic techniques for data collection largely inspired by data collection in psychological and psycholinguistic research – Munsell colour chips, salience tests, number of connotations per colour term, frequency in texts etc. (cf. MacLaury 2001 for an overview and references).

The major part of the data underlying Andersen's (1978) and Brown's (2005a-b) studies of BODY seem to come from available dictionaries and word lists. A new step in the study of BODY is undertaken in Majid et al. (eds.) (2006), where the papers on

ten different languages are all based on a consistent application of multiple methods for data collection, such as collecting replies to a detailed questionnaire and drawing outlines of the various body-part terms on a picture of a human body.

Elicitation of verbal descriptions for visual stimuli underlies a large portion of cross-linguistic work on MOTION (the pictures in the Frog Story, initially coming from the cross-linguistic work on child-language acquisition by Berman, Slobin and their colleagues, Berman & Slobin 1994, or video-sequences depicting moving objects), on DIMENSIONAL ADJECTIVES (Lang 2001) and, most recently, on THE CUTTING-BREAKING domain (videoclips showing various types of material separation, Majid & Bowerman (eds.) (2007). The papers on AQUAMOTION in Maisak & Rakhilina (eds.) (2007) contain detailed descriptions of particular languages by language experts, who have conducted their own in-depth studies (e.g., using corpora and dictionary searches, work with informants, introspection, etc.), while at the same time following the common checklist, or guidelines.

Finally, comparison of parallel translations of one and the same text into different languages is becoming a popular tool in typology. Viberg has been carrying out “small-scale” lexical-typological (or, rather, contrastive) studies of several verbs using parallel corpora in a few European languages (e.g., Viberg 2002, 2005, 2006). Parallel texts in more than 100 languages (the Gospel according to Mark) underlie Wälchli’s “large-scale” lexical-typological studies of MOTION verbs (Wälchli 2006; Wälchli & Zúñiga 2006).

#### 4.4 Questions and generalizations

Domain-categorization lexical typology, as typology otherwise, is driven by the curiosity to understand what is variable and universal in a particular linguistic phenomenon. Of central concern here are therefore questions such as according to what parameters a specific phenomenon can vary across languages, in what patterns these parameters (co-)occur and what generalisations can be made about attested vs. possible patterns.

One possible result of cross-linguistic studies is, of course, a *classification* of the obtained data into patterns, or types (and, in some cases, the corresponding classification of the languages). While classifications of phenomena and languages are useful on their own, their value increases if they can be *related to other phenomena* – both linguistic and non-linguistic ones. In Ricca’s (1993) study of DEICTIC verbs, the twenty languages are classified into three groups depending on the extent to which they make a systematic distinction between verbs showing centripetal (to the deictic centre, normally the speaker) vs. centrifugal (from the deictic centre) motion. Interestingly, the distribution of the types across the sample languages is dependent on a combination of genetic and areal factors and can therefore contribute to our general

understanding of genetically stable vs. borrowable (or diffusible) phenomena with further implications for research in historical linguistics. Thus, the *fully deictic* languages are mainly found in Southwestern and Southern Europe (Portuguese, Spanish, Italian, Albanian, Modern Greek, with the two Finno-Ugric outliers Hungarian and Finnish), the *non-deictic* languages are Western and Eastern Slavic and Baltic, while the *predominantly deictic* ones are Germanic, French and the two Southern Slavic languages Serbo-Croatian and Slovenian. The next logical step to be taken from there would be to provide both a better coverage of the European language varieties and a broader non-European sample. These will together contribute to a better understanding of the relative contribution of universal, genetic and areal factors in this domain (cf. Wilkins & Hill 1995 for similar considerations).

Brown’s (2005a-b) BODY-PART chapters in the *World Atlas of Language Structures* are further examples of classifications with interesting implications. First, they show different statistical asymmetries in the distribution of the types: while only 12% of the sample’s languages (72 languages of the 593 languages) use the same word for FINGER and HAND, the corresponding share for the non-distinction of HAND and ARM is much higher – approximately 37% languages in the sample (228 languages of the 617 languages). Second, the chapters show that the skewings in the areal distribution of the languages in each category are not random, but correlate either with geography (languages without HAND-ARM distinction tend to occur more frequently near the equator) or with culture (languages without FINGER-HAND distinction tend to be spoken by traditional hunter-gatherers or by groups having a mixed economy of cultivation and foraging). Brown suggests that the former can be dependent on the local clothing traditions (extensive clothing, sometimes also including gloves or mittens, greatly increases the distinctiveness of arm parts), whereas the latter can perhaps be explained by the greater use of finger rings among agricultural people which, in turn, makes fingers salient as distinct hand parts. In both cases, Brown’s explanation makes therefore appeal to the cultural practices. It should be remembered that the correlations are far from perfect: e.g., although Russian is being spoken in a much colder climate than Italian (and Russians are dependent on gloves and mittens to a much higher extent than Italians), the former uses one and the same word for HAND and ARM, while the latter has two.

Some classifications might seem to play a more central, essential role in the linguistic system, they can thus be used as *predictors* for other linguistic phenomena and might have important implications outside of the language. Consider the impact of Talmy’s (1985, 1991) studies on the research on MOTION verbs. Talmy focuses on the different “conflations” of the motion meaning component with other meaning components, or different lexicalization patterns, where lexicalization is defined to be “involved where a particular meaning component is found to be in regular association with a particular morpheme” (Talmy 1985: 59). It is generally assumed that languages

tend to be consistent with respect to their lexicalization patterns in motion verbs (cf. the classification into *path-conflating* vs. *manner-conflating* vs. *figure-conflating* languages in Talmy 1985, its later modifications into *satellite-framed* vs. *verb-framed* languages in Talmy 1991 and even more radical reorganizations in Slobin & Hoiting 1994 and in Croft 2003: 222). The assumption is, thus, that lexicalization in this domain is not an issue of one particular meaning (or one particular combination of meaning components) being expressed via a word, but is more *global* – i.e., pertaining to the whole MOTION domain or, at least, to a major part of it. There is also ample research on possible connections between the lexicalization type of a certain language and, say, its discourse organization, child language acquisition of the domain, even non-verbal communication such as gestures and “thinking for speaking”, i.e., mild cognitive effects of linguistic relativity (e.g., Slobin 2003; Slobin & Bowerman 2007; Kita & Özyürek 2003). In short, a language’s lexicalization pattern in the MOTION domain is often taken to belong to its *typologically relevant properties*.

However, the role of the “Lexicalization Pattern” theory should not be exaggerated, since it is based on a limited number of languages, where, in addition, only a subset of motion verbs has been studied in detail. Wälchli’s (2006, 2006/7) on-going research on MOTION events based on parallel texts in more than 100 languages has already challenged some of the basic assumptions and predictions of the “Lexicalization Pattern” Theory. For instance, languages with a consistent (or predominantly consistent) “global” lexicalization-pattern type seem to be fairly unusual – the best examples are in fact found among the Eurasian and North American languages (exactly those that have figured prominently in Talmy’s seminal papers). Many more languages, however, turn out to show mixed behaviour. And in fact, even languages that are considered to fit into the Talmy-Slobin-Croft typology, often have verbs showing a deviant behaviour. For instance, the *Path-conflating* verbs ‘fall’ or ‘sink’ are present in many otherwise *Manner-conflating* languages like English and Swedish, while the otherwise *Path-conflating* language French has a general motion verb *aller* which does not conflate with any direction (Viberg 2006: 113). There are, therefore, asymmetries even among the verbs belonging to the same domain, which may be ordered rather than being random. As Viberg hypothesizes, *Path-conflation* might be most frequent with verbs for UNCONTROLLED MOTION (DOWN), least frequent with verbs for GENERAL DIRECTION (TO/FROM) and intermediate with verbs for CONTROLLABLE MOTION (UP/DOWN, IN/OUT). These can be said to show a markedness hierarchy – the notion we will discuss shortly.

Classifications become particularly interesting when they show various *asymmetries* – statistical preferences for certain combinations of parameters and the absence of attested though logically possible types. Consider Nerlove and Romney’s (1967) SIBLING terminologies. These are based on eight logical KIN types as defined by three

parameters (sex of ego, sex of sibling, relative age) – e.g., whether one and the same term is used for all siblings, whether there are two separate terms for ‘brother’ and ‘sister’, whether there are four different terms (‘younger brother’, ‘elder brother’, ‘younger sister’, ‘elder sister’), etc., which leads to 4140 logically possible types. However, only 10 of those are attested in more than one language in the 245-language sample! In a similar vein, even though, perhaps, less spectacularly, one of the five logically possible types is not attested in the KINSHIP typologies focusing on the patterning of terms for PARENT/UNCLE/AUNT (with the roots in Morgan 1870 and further modified by later research). Thus, languages can have the same term for ‘father’, ‘father’s brother’ and ‘mother’s brother’ (the “Hawaiian”, or “Generational” type), or the same term for ‘father’ and ‘father’s brother’, as opposed to ‘mother’s brother’ (the “Iroquois”, or “Bifurcate merging” type). However, no language has so far shown a system with one and the same term used for ‘father’ and ‘mother’s brother’, to the exclusion of ‘father’s brother’.

The two kinship asymmetries cited above are examples of *linguistic universals* – generalizations on what is generally preferred/dispreferred (or even possible/impossible) in human languages – the search for which has been a high priority on the typologists’ agenda for a long time. Linguistic universals within the lexical-typological research are often formulated in terms of *lexicalization hierarchies/implications* and *lexical universals*. The task is then both to unveil and to explain them. Continuing on the issue of the Morgan-inspired KINSHIP terminologies – why is one of the five logically possible lexicalizations not attested? It has been suggested that the reason is cognitive: a definition of a term covering both ‘father’ and ‘mother’s brother’ would be cognitively more complex than the other four lexicalizations, since it will require disjunction (‘father’ or ‘mother’s brother’, cf. ‘male relative of one’s patriline’ for ‘father’ and ‘father’s brother’). It has also been suggested that the PARENT/AUNT/UNCLE terminology has the central role in predicting other properties in the KINSHIP systems, such as, e.g., the SIBLING terminology (where the generalizations can be formulated as lexicalization implications); this, in turn, can be explained by the role of the corresponding relations in regulating marriage possibilities (for details cf. Evans 2001).

Most lexicalization generalizations build on the unequal status of different words and other lexicalized expressions – either for encoding a particular conceptual domain, or in general. Some words are *basic*, while others are *derived*, some are *less marked*, whereas others are *more marked* – defined by various criteria stemming from Berlin & Kay’s (1969) classical study of colour terms, and the numerous publications by Greenberg (e.g., 1966), Brown (e.g., 1976) and by Berlin (e.g., 1992). One of Greenberg’s own examples nicely links to the discussion of kinship terminology in the preceding paragraph. Greenberg (1980) suggests that each of the different semantic

components used for analyzing kin terms defines its own markedness hierarchy, as shown in Table 4:

**Table 4.** Markedness relations among kin terms according to Greenberg (1980)

Less marked	More marked	Example
LINEAL	COLLATERAL	'father' < 'uncle'
CONSANGUINEAL	AFFINAL	'brother' < 'brother-in-law'
LESS REMOTE (measured in number of generations)	MORE REMOTE (measured in number of generations)	'father' < 'grandfather'
SENIOR (including both seniority within one generation and the distinction ascending-descending)	JUNIOR	'older brother' < 'younger brother'

Markedness, according to Greenberg, would manifest itself in

- zero expression vs. overt expression for certain categories cf. the CONSANGUINEAL vs. the AFFINAL relation in such pairs as *brother* vs. *brother-in-law*;
- defectivation, i.e., the absence of a term in the marked category which would correspond to an existing one in the unmarked category cf. \**cousin-in-law*;
- neutralization of certain distinctions in some categories (cf. neutralization of sex reference in *cousin* as against *brother* and *sister*), and
- higher text frequencies for unmarked categories.

When all the individual markedness hierarchies are combined the least marked of all the kin terms turn out to be PARENTAL TERMS, which is further corroborated by various properties singling out these terms across languages (cf. Dahl & Koptjevskaja-Tamm 2001).

Since the discussions of basic vs. non-basic terms in the domain of COLOUR are by now well known and widely quoted, the reader is referred to the relevant literature. Let us look at a couple of other examples of lexicalization generalizations pertaining to PERCEPTION verbs and BODY-PART terminologies.

Viberg (1984, 2001) suggests that PERCEPTION verbs across languages follow the sense-modality hierarchy

sight > hearing > touch >  $\left\{ \begin{array}{l} \text{smell} \\ \text{taste} \end{array} \right.$

Not only does the hierarchy conform to the standard markedness criteria, but it also restricts patterns of intrafield polysemy, or semantic extensions of perception verbs within the PERCEPTION domain: for instance, in Russian the verb *slyšat'* "to hear" is often used in combination with the noun *zapax* "smell" for reference to smelling (a verb normally relating to a higher sense modality extends its uses to lower modalities,

whereas the opposite does not seem to be attested) (cf. also Vanhove this volume; for a possible exception cf. Maslova 2004).

There is a relatively long tradition of cross-linguistic markedness generalizations about the inventories of BODY-PART terms, starting with Brown (1976) and Andersen (1978) (and further developed in Brown 2001; Wilkins 1996), e.g.:

- If both HAND and FOOT are labelled, they are labelled differently (cf. English and Italian in Table 2);
- If there is a distinct term for FOOT, then there will be a distinct term for HAND (cf. English, Italian, which have both, vs. Japanese and Rumanian, which have a distinct term for HAND, but not for FOOT, and Russian, that lacks either in Table 2).

BODY-PARTS differ, obviously, in how they relate to each other and to the whole body; the nose is, for instance, a part of the face, which, in turn, is a part of the head, which, in turn, is part of the body. BODY-PART terminology is therefore interesting for cross-linguistic generalizations on the partonomic levels of depth, or ethnolinguistic partonomy. Andersen (1978) suggests, for instance, that

- there are never more than six levels of depth in the partonomy relating to BODY PART terminology, and
- there will be distinct terms for BODY, HEAD, ARM, EYES, NOSE and MOUTH

The ten language descriptions in Majid et al. (2006) challenge a high portion of the earlier cross-linguistic research on categorization and linguistic/conceptual segmentation of the BODY. Thus, Lavukaleve, a Papuan language isolate spoken on the Russell islands within the central Solomon islands (Terrill 2006: 316) has one and the same word, *tau*, for both ARM and LEG, contradicting the claim that ARM is always lexicalized by a distinct term. In addition, Lavukaleve has a distinct simple word, *fe*, for reference to FOOT, but nothing comparable for HAND – contradicting therefore another of the claims above. Lavukaleve (as well as some of the other languages in the volume, e.g., Savosavo, another Papuan language in the Solomon islands, cf. Wegener 2006) shows also that linguistic evidence for partonomic relations between subset of body part terms can be difficult to achieve and that languages do not necessarily show any multi-level conceptualization in this domain, thus contradicting some of the universals proposed in Brown (1976) and Andersen (1978).

#### 4.5 Explanations

Linguistic typology in general seeks explanations for two different kinds of observed facts and generalizations – first, for the observed patterns in the linguistic phenomena themselves and, second, in their distribution across languages. A few examples of the

latter have been given above (e.g., in connection with DEICTIC/NON-DEICTIC verbs and the HAND/ARM vs. FINGER/HAND distinctions).

Examples of the former are, for instance, the standard explanations for the possible COLOUR-TERM systems as governed by the neurophysiology of vision (Berlin & Kay 1969; Kay & McDaniel 1978; Kay & Maffi 2000). Vision itself is also dominant among the sense modalities, as is well established within cognitive psychology and neurophysiology, which explains its highest position in the hierarchy for PERCEPTION verbs (Viberg 2001: 1306–1307). Biology-rooted factors (e.g., perceptual discontinuity and other properties derived from perception), as well as functions have also been suggested as underlying segmentation of the BODY across languages (Brown 1976; Andersen 1978; also Enfield et al. 2006 for the discussion of the explanations and some counterexamples). The unmarked status of PARENTAL TERMS in the kin term systems follows, of course, from the unique biology-rooted status of parents (or at least of the biological mother) in any person's life. All these examples, together with the earlier cognitive-based account for the cross-linguistically unattested lexicalization of 'father + mother's brother' to the exclusion of 'father's brother, suggest that lexical categories can be motivated – at least partly – by non-linguistic cognition and shaped by human perceptual and cognitive predispositions.

There are also other possible explanations for cross-linguistic lexical preferences. We have already seen some examples of *social and cultural* practices as explanations for the cross-linguistically recurrent lexicalization patterns in connection with HAND/ARM, FINGER/HAND and PARENT/UNCLE/AUNT KINSHIP terminologies. Now, consider the above-mentioned cross-linguistic preference for conflation of PATH (DOWN) and UNCONTROLLED MOTION in verbs like *fall* or *sink* (Viberg 2006: 113). An obvious explanation here lies in the power of the omnipresent gravity in human environment: uncontrolled – and often unintended motion – under normal conditions in the default case means falling or sinking (in water). The long tradition of research on kinship has on the whole been an arena for hot disputes on the role of universal (biology-rooted) vs. social-construction explanations (see Foley 1997: 131–149 for an overview). Numerous examples of *environmental*, *social* and *cultural* explanations, combined with biology-rooted factors are found in the recent research on colour (e.g., Wierzbicka 1990; MacLaury 2001; Jameson 2005; Dedrick 2005; Paramei 2005 and the references there).

Other possible explanations for cross-linguistically recurrent lexicalizations include the natural logic of events (rational and purposive connections among the components of the same event, e.g., Enfield 2007) and – last, but not least – innate concepts.

#### 4.6 Universal vs. language-specific lexicalizations?

Let's come back to the initial questions asked in the beginning of Section 4. What meanings can be or cannot be expressed by a single word in different languages, or what word meanings are universal, frequent, possible, impossible? Are there any uni-

versal (or at least statistically predominant) restrictions on the meanings that can or cannot be expressed by single words across languages, or are languages more or less free to choose here?

It is in fact difficult to evaluate to what extent these questions have been approached and answered in the research up to now. Apart from the fact that very few concepts and conceptual domains have been studied at all, the different research traditions are hard to bring into line with each other, their results are often incommensurable and hardly mutually translatable. Take the hierarchy of perception verbs mentioned above. What does it actually mean for the universal-lexicalization enterprise? Does it imply that some languages will not lexicalize the meaning 'to hear' and many more languages will not lexicalize the meanings "to taste", "to touch" and "to smell"? Conversely, if this is true (or at least partially true) – wouldn't that imply that the word referring to hearing in a language which also has designated expressions for tasting, smelling and touching, will have a different meaning from the word that covers perception in all these modalities? Likewise, would the existence of the languages that merge 'father' and 'father's brother', as well as 'mother' and 'mother's sister' mean that biological parents do not correspond to universally lexicalized meanings? And how does this relate, in turn, to Greenberg's markedness universals for kin-terms and the "parental prototype" suggested both there and in Dahl & Koptjevskaja-Tamm (2001)?

As becomes clear, the answers to these questions are crucially dependent on what stance we take on the issues of polysemy and approximate vs. precise meaning identity discussed in Section 3. The overview in Goddard (2001; cf. also Brown 2001) which makes a serious attempt to take polysemy into consideration and aim at precise meaning definitions (in the NSM tradition), is probably the most updated one over suggested lexico-semantic universals; the important predecessors to the paper include the collective volume (Goddard & Wierzbicka 1994). As Goddard argues, "see" and "hear" seem to stand the proof of being universally lexicalized (at least as separate meanings within polysemous expressions). Some presumably basic and universal "concepts" seem to be doubtful as lexical universals or, at least, can only be viewed as approximate ones (e.g., "eat", "give"). A few of the other surprises include the non-universal status of "water" and "sun" based on the fact that languages can have more than one word for each of those (cf. "hot water" vs. "non-hot water in Japanese, "sun low in the sky" vs. "hot sun overhead" in Nyawaygi, Australia). Emotions, contrary to many common assumptions, turn out to be highly culture- and language-specific. "Mother", in its biological sense, has reasonable chances to survive as a lexical universal too, whereas "father's" chances are considerably lower (since his role is much more subject to social factors). This, in fact, leads to interesting asymmetries in kinship terms, where "mother + mother's sister" is supposed to be polysemous, whereas "father + father's brother" is "allowed" to be semantically general – a fact not mentioned by Goddard (2001). All in all, according to Goddard, the best candidates for the universally lexicalized meanings

turn out to be overwhelmingly found within the set of semantic primitives suggested within the NSM (which now covers about sixty items).

However, opinions vary as to what can count as universal lexicalizations and on the kinds of evidence for or against them (cf. Brown 2001), also among the individual researchers who have contributed to the papers in Goddard & Wierzbicka (1994). Consider ‘want’, which the NSM considers a semantic primitive, lexicalized in all languages. A recent study casting doubts on the suggested universality of ‘want’ is Khanina (forthc.), who shows that language after language in a sample of 73 genetically, areally and structurally diverse languages merge this meaning with other meanings (often modal and mental-emotive) in one and the same lexeme. In fact, exclusively desiderative expressions are predominantly found in the languages of Eurasia, excluding South Asia, and those of Northern America, while two thirds of the desiderative expressions in Khanina’s sample show other meanings as well. In many of these, a case can be made for polysemy; in others, however, there do not seem to be obvious morphosyntactic differences between desiderative and other uses. This does not exclude that further analysis will not provide arguments for polysemy even there. However, the fact itself that ‘want’ relatively seldom has a “lexeme of its own”, but tends rather to share the same lexeme with other meanings, is significant and has to be taken into consideration in discussions of universally lexicalized meanings. Khanina’s own conclusion is that the status of ‘want’ is subject to cross-linguistic and cross-cultural variation, as many other concepts: in some cases it is an indivisible and indefinable salient concept of the culture, whereas in others it is not salient, but is treated only as particular type of a more general situation.

It is not always clear how precise semantic identity is established cross-linguistically and to what extent it is interesting and important to achieve that degree of exactness. Many scholars take cross-linguistic semantic comparability fairly easily, without always being aware of this. This is, for instance, characteristic of the research areas presented in the next two sections. In certain cases this is undoubtedly justified by the task and aims of the research whereby a higher degree of semantic precision might render the task undoable and create obstacles for interesting generalizations; in other cases, on the contrary, semantic vagueness cannot be justified and can even be shown to be detrimental for an effective cross-linguistic comparison.

##### 5. What different meanings can be expressed by one and the same lexeme or by words derivationally related to each other?

What different meanings can be expressed by one and the same lexeme, by lexemes within one and the same synchronic word family (words linked by derivational relations) or by lexemes historically derived from each other? These questions can be

approached from different angles. Most of the other contributions to the volume start with an *individual lexical item*, or with several items belonging to one and the same *individual lexical field*, and ask what other lexical or grammatical meanings can be expressed by the same form(s) or by forms derived from it/them. In the sub-sections below we focus on semantic relations between particular lexical units, or particular meanings – i.e., particular instances of *semantic motivation*.

But we can also compare *whole classes or groups of words* where one of the classes contains words derived from, or formed on the words in the other one, and ask about the semantic relations associated with a particular word formation device. The focus here is on the *regularities* in lexical motivation seen as an interaction of formal (morphological) and semantic motivation (cf. Koch & Marzo 2007).

The next two subsections will give a short overview of the current cross-linguistic research along the two lines. The bulk of examples come from the BODY domain, but we will also briefly touch upon some studies on derivational morphemes and compounding.

##### 5.1 Focusing on semantic motivation: Another look at the BODY and outside

When *foot* and *head* are used with meanings different from their normal body-part meanings in the expressions *the foot of the mountain* and *the head of the department*, they present clear cases of *polysemy*. In other cases two meanings do not coexist within one and the same lexeme, but are related diachronically. The Italian *testa* “head” has developed from the Latin word *testa* “splinter”, but does not show the original meaning any longer. On the other hand, such meanings of the German word *Haupt* as “head (of the department, delegation etc.)” and “top (of a mountain)” have developed from its earlier BODY-PART meaning “head”, for which *Haupt* in Modern German has been more or less replaced by *Kopf*. A useful cover term for all these cases is *semantic shift*, which refers to a pair of meanings A and B linked by some genetic relation, either synchronically or diachronically. Diachronic semantic shifts can sometimes lead to *heterosemy* (rather than to polysemy), which refers to “cases (within a single language) where two or more meanings or functions that are historically related, in the sense of deriving from the same ultimate source, are borne by reflexes of the common source element that belong in different morphosyntactic categories” (Lichtenberk 1991: 476). The pair *head* (like *my head*) and *ahead* (*ahead of me*) is an example of heterosemy.

It is useful to distinguish between *intrafield semantic shifts*, that relate two meanings belonging to the same semantic domain, and *interfield*, or *transfield semantic shifts*, that relate meanings belonging to different semantic domains (all those quoted above). What might count as intrafield polysemy of body-part terms is a difficult question which brings to the fore the problem of distinguishing semantically general meanings,

on the one hand, from polysemy – cf. the discussion in sections 3.2. and 4. It has, for instance, been widely debated whether *ruka* “hand/arm” in Russian is semantically general or polysemous (cf. Wierzbicka 2007 for a recent hefty argumentation in favour of polysemy); it might very well turn out that a large portion (or even all) of the languages classified as neutralizing the HAND/ARM or the FINGER/HAND distinctions in Brown (2005a,b) show in fact recurrent patterns of intrafield polysemy rather than semantic generality. I will not dwell on intrafield polysemy or on derivation of body-part terms in this paper. There has been interesting cross-linguistic research on diachronic semantic shifts leading to the emergence of various body-part terms (both intra- and interfield shifts) – cf. Koch (this volume) and the references there.

The best cross-linguistically studied cases of semantic associations pertaining to body involve body-part terms as *grammaticalization sources* for markers of SPATIAL RELATIONS and REFLEXIVE-MIDDLE-RECIPROCAL markers (in most cases leading to heterosemy).

Various cross-linguistically interesting generalizations on grammaticalization BODY-PARTS → SPATIAL RELATIONS and explanations for them are found in Svorou’s (1993) cross-linguistic study covering 55 languages of the world, and Heine’s (1989) and Bowden’s (1992) studies of the numerous African vs. Oceanic languages. Thus, FRONT-REGION relations are often expressed by markers coming from terms for eye, face, forehead, mouth, head and breast/chest, while those for back-region relations often emerge from body-part terms for back, buttocks, anus and loins (Svorou 1993: 71–72), e.g.,

- (2) Examples of development BODY-PARTS → SPATIAL RELATIONS  
(Svorou 1993: 71–72)
- a. Halia (Austronesia, Oceanic, NW and Central Solomons)  
*i matana* “in front of” < *i* “in, at” + *mata* “eye” + *-na* (ADV.SUF)
  - b. !Kung (Khoisan)  
*tsi’i* “in front of” < *ts’i* “mouth”
  - c. Navajo (Na-Dene)  
*bi-tsi* “in front, at” < *’atsii* “head, hair”
  - d. Basque (Isolate)  
*gibelean* “in back of” < *gibel* “back” + *-ean* (LOC)
  - e. Papago (Aztec-Tanoan)  
*-’ai* “in back of” < *’aat* “anus”

There are significant cross-linguistic differences here: terms referring to one and the same body-part can sometimes give rise to different developments; in addition, languages can also differ in their preferences for using particular body-part terms as sources for spatial markers. For instance, in Svorou’s sample, ‘head’ gives rise to markers of BACK-REGION in 12 cases and to markers of FRONT-REGION in 2 cases, while ‘back’ gives

rise to markers of BACK-REGION in 15 cases, of TOP-REGION in 3 cases and of BOTTOM-REGION in 1 case. There are interesting explanations for these facts, based on *universal preferences* and on *areal/genetic* factors. Some of the differences in the semantic developments of “comparable” body-part terms have been interpreted as consequences of two different models, according to which anatomy is mapped into spatial relations – the *anthropomorphic* model (corresponding to the canonical upright position of a standing man with his/her back oriented backwards), and the *zoomorphic* model (corresponding to the canonical position of an animal standing on its four legs with its back oriented upwards) (Svorou 1993: 74–76; Heine 1997: 37–49). The anthropomorphic model is cross-linguistically preferred in being found more often across languages; in addition, even languages with predominantly zoomorphically modelled spatial concepts have at least some that emerge from the human body (Svorou 1993: 72–75; Heine 1997: 40). This is, of course, not particularly surprising given the general human predilection for anthropocentricity and embodiment in its different aspects. Zoomorphically modelled spatial concepts, on the contrary, seem to have a clearly areal and genetic distribution.

There are also remarkable “areal differences in the relative weight given to the three major body regions” (i.e., head, trunk, and extremities, Heine 1997: 43. The proportions among the relevant semantic shifts in the Oceanic languages (Bowden 1992) more or less correspond to those in Svorou’s global sample in that the ‘head’ and the ‘trunk’ each provide about 49% of the sources. The African languages (Heine 1989), on the other hand, significantly differ in the proportions between the ‘head’ (38%) and the ‘trunk’ (60%) with the ‘belly’ as the more prominent source for spatial orientation in Africa than elsewhere in the world.

The brief presentation above shows, thus, that the cross-linguistic studies on the grammaticalization path BODY-PARTS → SPATIAL MARKERS “live up” to such expectations of modern typological research as proposals of explanations for the cross-linguistic patterns and for the asymmetries in their distribution across a large sample of languages. Another relatively well-studied group of shifts with BODY-PARTS as source is involved in grammaticalisation of REFLEXIVE-RECIPROCAL-MIDDLE markers. Schladt’s (2000) cross-linguistic study based on 150 languages shows that ‘body’ constitutes the absolutely most frequent source for reflexive markers, while ‘head’ is one of the other major ones. Since REFLEXIVE markers in turn tend to develop further into RECIPROCAL and MIDDLE markers, ‘body’ and ‘head’ may also grammaticalize into those as well (cf. Heine & Kuteva 2002 for discussion and numerous examples). Although Schladt’s sample is geographically biased (and contains many more African languages than languages from the other parts of the world), it seems sufficient for manifesting interesting areal differences in the distribution of the grammaticalization sources. Thus, the reflexive markers in the African languages are derived much more frequently from ‘body’ and ‘body parts’ than elsewhere; in addition, within this category, ‘body’ is the much more preferred option

than ‘head’ in Africa than in Asia and Europe, whereas the languages of Northern America use exclusively ‘body’.

There are other well-known grammaticalization paths from body-part terms attested across languages, but studied in a less systematic cross-linguistic way, e.g., the following ones:

BODY-PARTS → NUMERALS, in particular, ‘hand’ → ‘five’ (e.g., *lima* “hand, five” in Samoan and all over Austronesian). As Heine & Kuteva (2002: 166) suggest, “[n]ouns for ‘hand’ probably provide the most widespread source for numerals for ‘five’ in the languages of the world”, but it is not quite clear what sample underlies this generalization. ‘Finger’ sometimes occurs in expressions for ‘six’, such as ‘a finger passes hand’, while ‘foot’ may occur in expressions like ‘two hands, one foot and one finger’ for ‘sixteen’ (Harald Hammarström p.c.);

‘hand’ → POSSESSION, for which Heine & Kuteva (2002: 167) give a few examples from African languages and suggest that this is an areally induced process. However, Estonian demonstrates a similar grammaticalization path (Ojutkangas 2000) – incidentally, several body-part terms in Estonian have also grammaticalized into spatial postpositions (related phenomena have a wider attestation in Finnic and even Finno-Ugric, Bernhard Wälchli p.c.).

It is rather striking that none of the numerous cross-domain semantic extensions from body-part terms not leading to grammaticalization and attested all over the world, has been studied in a systematic cross-linguistic way, at least remotely comparable to the studies reported on above. A particularly interesting topic is the use of BODY-PART terms in conventionalised descriptions of EMOTIONS and MENTAL STATES, a phenomenon found all over the world, e.g.,

- (3) Lao (Tai-Kadai) (Enfield 2002: 87)
- |    |                         |    |                      |    |                      |
|----|-------------------------|----|----------------------|----|----------------------|
| a. | <i>aj3-hòdn4</i>        | b. | <i>caj3-dii3</i>     | c. | <i>caj3-kaa4</i>     |
|    | heart-hot               |    | heart-good           |    | heart-daring         |
|    | ‘impatient, hot-headed’ |    | ‘nice, good-hearted’ |    | ‘daring, courageous’ |
- (4) Kuot (isolate, “Papuan”, New Ireland, Papua New Guinea) (Lindström 2002: 162)
- |   |               |           |
|---|---------------|-----------|
| <i>gigina-m</i>                                   | <i>[dal’p</i> | <i>a]</i> |
| heavy-3PL   | stomach.NSG   | 3M.POSS   |
| ‘he is worried/sad’ (lit. ‘his stomach is heavy’) |               |           |
- (5) Yéli Dnye (isolate, “Papuan”, Rossel island, Papua New Guinea) (Levinson 2006: 237)
- |   |                   |
|---|-------------------|
| <i>a nuu u</i>  | <i>tpile</i>      |
| my throat   | his/its/her thing |
| ‘A thing I really like it’ (lit. ‘My throat its thing’) |                   |

As the examples above demonstrate, languages can differ significantly in which body-parts can be seats for which emotions: the majority of emotional descriptions

in Lao, Kuot and Yéli are based on ‘heart’ vs. ‘stomach’ vs. ‘throat’. Some areal tendencies have been suggested: thus, Lao seems to be representative for the languages in Southeast Asia (Matisoff 1986), which often use the word for ‘heart’ or ‘liver’ in most emotion descriptions. Enfield & Wierzbicka (2002) contain a number of enlightening papers on different languages, and a first important step towards a more systematic cross-linguistic comparison is provided by the volume on the role of heart in expressions of emotions (Sharifian et al. eds. forthc.). Large-scale cross-linguistic comparisons in this area are, however, still lacking – and are badly needed, in particular, given the prominent role of both body and emotions in the current cognitively-oriented semantic theories. The methodological and theoretical problems for such comparisons are, however, overwhelming. On the one hand, as we have seen, categorization of the body itself is subject to significant cross-linguistic variation. On the other hand, unveiling and describing the meaning of emotion expressions even in one’s native language is a much more difficult enterprise than in many other domains, emotions are to a high degree culture-specific, and systematic cross-linguistic comparison of emotion expressions has hardly begun (the very interesting papers in Harkins & Wierzbicka (2001) contain descriptions that are not directly comparable to each other).

Some types of semantic associations involving body-part terms are probably restricted to certain linguistic areas or to particular linguistic families. Among the recurrent features of Australian lexical systems Evans (1992: 479) mentions examples of “synecdoche, by means of which animals or plants are named for their most salient body-part”. Thus, ‘tooth’ is extensively used in such names, leading to the polysemy of *muyiny* in Wadyiginy “dog, wild asparagus”, or to the different meanings within “the set of cognates of \**waartu* including Umbugarla *waartu* “mosquito”, Yolngu *wart:u* “dog”, Kayardild *wardu* “sandfly” and *wardunda* “mangrove rat””. The typical recurrent Australian metaphors include extensions of ‘eye’ to any point-like entities, including ‘star’, ‘well’, ‘small hole in ground’ and ‘bullet, and the use of ‘ear’ as the seat of intelligence and apprehension, e.g., ‘ear-bad’ or ‘ear-without’ = ‘crazy’, etc.

The cross-linguistic research on semantic associations and semantic shifts in domains other than BODY is even less systematic. The best-studied cases involve again heterosemy and grammaticalization from different sources (e.g., motion and posture verbs, “give”, “acquire”, etc.; cf. also the discussion of “want” in Section 3.6. – cf. Heine & Kuteva 2002 for the numerous examples and references). One of the notable few exceptions is the research on perception verbs developing cognitive meanings (Sweetser 1990; Evans & Wilkins 2001, cf. also Vanhove this volume). Brown (2001) reports on several other cross-linguistically recurrent connections (‘wood’ vs. ‘tree’, ‘seed’ vs. ‘fruit’, ‘wind’ vs. ‘air’), for which social and cultural factors have been suggested. Thus, speakers of languages whether there is polysemy between “wood” and “tree” (two thirds of the languages in a big sample) usually live in small-scale, traditional

societies, while speakers of languages separating them usually live in large nation states (cf. also Section 4.3 for the potential BODY-PART examples).

## 5.2 Formal motivation and its semantic correlates

The preceding section has dealt with cross-linguistic studies on *particular* instances of semantic motivation. In this section we will briefly touch on the issue of the *regularities* in lexical motivation seen as an interaction of formal (morphological) and semantic motivation (cf. Koch & Marzo 2007). The issues we discuss here include the following:

- what meanings correspond to “more basic” vs. “regularly derived” words?,
- what formal (primarily morphological) devices are there in a language for forming words from other words, or lexical units from other lexical units – e.g., derivation, compounding?, and
- what meaning relations can be expressed by these devices?

Three examples of recent large-scale typological studies will give a flavour of how these questions can be approached and answered. The first of them, Nichols et al. (2004) approaches the issue of *more basic* vs. *regularly derived* words and how it interacts with the *formal word-forming devices* in a language. Given two sets of words such that the words in one set are semantically almost identical to those in the other apart from one and the same meaning component (or, put differently, the words in the two sets are related to each other by means of one and the same meaning relation), are there systematic formal relations between the words in the two sets and if yes, which of the words will be formally more basic vs. derived from the others? The two other studies, Juravsky (1996) and Wälchli (2005), focus on a particular word-forming device, derivation of diminutives and co-compounding, and ask the third of the questions stated above, namely, *what meaning relations can be expressed by them*. Although both studies concentrate on languages that have these devices, Wälchli (2005) is also interested in the question of absence vs. presence and, generally, of uneven distribution of co-compounding across languages, i.e., in the second of the above stated questions.

The first example of large-scale cross-linguistic study that takes regularities of lexical motivation seriously is an important recent contribution by Nichols et al. (2004). It presents a typology of 80 languages based on their treatment of what the authors view as semantically basic and almost universal intransitive verbs such as ‘sit’, ‘fear’, ‘laugh’, ‘fall’ and their transitive counterparts (all in all 18 pairs). The list of verb pairs takes into consideration various parameters that are known or supposed to have an impact on derivational processes (animacy, agency, resistance to force), to be sufficiently common and easily found in lexical sources and/or translatable, and to show a wide spread in lexical semantics.

The main question is whether the two sets of words are formally related to each other and if yes, how – i.e., which of the classes contains words derived from the words in the other one. It turns out that languages tend to be consistent in whether they treat intransitives as basic and transitives as derived by means of causative morphology (*transitivizing* languages), whether they derive intransitives by means of anti-causative morphology (*detransitivizing* languages), whether both intransitives and transitives are encoded by the same labile verb (*neutral* languages) or whether both intransitives and transitives have the same status (*indeterminate* languages), cf. Table 5.

**Table 5.** Examples of the types of lexical valence orientation illustrated by the derivational relations between the intransitive and transitive ‘hide’ (based on Nichols et al. 2004).

	<b>Chechen: transitivizing</b>	<b>Russian: detransitivizing</b>	<b>Thai: neutral</b>	<b>Nanai: indeterminate</b>
‘hide’ (go into hiding)	<i>dwa+lehq’</i>	<i>prjatat’-sja</i>	<i>àop</i>	<i>siri-</i>
‘hide’ (put into hiding)	<i>dwa+lehq’a-d-</i>	<i>prjatat’</i>	<i>àop</i>	<i>djaja-</i>

The study is, thus, based on a principled sample of lexemes, and the resulting language types stem from their frequencies in the principled global sample of languages – which, as the authors themselves claim, defines this work as piece of lexical typology. There are also further various statistically significant generalizations on the “inner logics” of the types themselves, on their correlation to other linguistic phenomena (grammatical and lexical, e.g., alignment and voice alternations, complexity, aspect and Aktionsart) and on their distribution across the languages of the world – in the robust tradition of the standard modern large-scale typological research.

Nichols et al. (2004) builds on a long tradition of cross-linguistic and typological studies on causatives, anti-causatives and, more generally, transitivity alternations (the cumulative results of which could be profitably used for constructing the list of verb pairs) – which does not reduce the groundbreaking character and the value of the study itself. Juravsky (1996) and Wälchli (2005), each focusing on one particular word-forming (derivational vs. compounding) device across a large number of languages, are rather of a more exploratory nature, even if they both have theoretical and methodological predecessors (partly coming from the same tradition as Nichols et al. 2004; e.g., Kemmer 1993; Haspelmath 1987, 1993).

Juravsky (1996) focuses on one particular derivational category – *diminutives* – identified via their ability to mean at least ‘small’. Wälchli (2005) studies *co-compounds* (dvandva compounds, pair words, copulative compounds) – compounds like ‘father-mother’ for parents, ‘milk-butter’ for ‘diary products’, etc., which are opposed to *subordinate compounds*, like ‘fingertip’ or ‘apple tree’, in which there is an asymmetric relation between the two parts (the head and its modifier). Each of these categories can be expressed by a variety of morphological devices – e.g., affixation, shifts in consonants,

vowels or lexical tones, and changes in noun-class or gender for diminutives. Each of them has also a number of different meanings and uses and, in addition, is not universal at all, even though both diminutives and co-compounds exist in many languages.

It is instructive to compare how the two studies approach their object. Juravsky's main interest lies in the amazing variety of semantic functions expressed by diminutives, in addition to the meaning 'small', e.g., 'child/offspring' (Tibetan *dom* "bear" vs. *dom-bu* "bear cub"), small-type (Ewe *hē* "knife" vs. *hē-vi* "razor"), "imitation" (Hungarian *csillag* "star" vs. *csillagocska* "asterisk"), intensity/exactness (Latin *parvus* "small" vs. *parvulus* "very small"), approximation (Greek *ksinos* "sour" vs. *ksinutikos* "sourish") and individuation (Yiddish *der zamd* "sand" vs. *dos zemdl* "grain of sand, Juravsky 1996: 536). They are also known to easily acquire connotations (or uses) of affection, sympathy and endearment or, conversely, of contempt, and are extensively used for various pragmatic functions, such as politeness. The main question is then what other meanings and uses can be attributed to the diminutive marker and what is the rationale behind this. To account for this, Jurafsky proposes a structural polysemy model (inspired by Lakoff's radial-category notion) in which the different senses displayed by diminutives are modelled together with the metaphorical and inferential relations among them. The model has both synchronic and diachronic applications, where the latter cover, among others, possible lexical sources for the category itself (words semantically or pragmatically linked to children).

Juravsky's study leaves many questions. A major problem is the lack of "standard typological" systematicity in the account for the data and for the analysis, even though the study is based on extant grammars and work with consultants for more than 50 genetically, structurally and areally diverse languages. Thus, the different meanings and uses underlying the radial category are merely presented by illustrations from one or several languages, without any overview over their distribution. We do not know which of the functions are cross-linguistically more or less common, and what kinds of genetic, areal or other patterns there are in the distribution of diminutives and of their various meanings and uses. Grandi (2002) shows, for instance, that diminutives and augmentatives are an interesting areal phenomenon in the Mediterranean languages, with some properties distinguishing them from the other genetically related languages.

The rationale behind Jurafsky's semantic map is not quite clear either, as opposed to the logic of semantic maps common in typological research (Haspelmath 2003, François this volume). All this makes the use of "universal" in the title somewhat doubtful, but compared to the majority of studies anchored in cognitive semantics with its inclination to allegedly universal claims, Jurafsky does provide ample cross-linguistic data and opens opportunities for future research.

Wälchli (2005) concentrates on several aspects of *co-compounding*. Co-compounds occur in several semantic types and functions, show a variety of formal patterns and considerable differences in frequencies of textual occurrences. Some

languages lack co-compounds, even those that have productive noun-noun compounding (e.g., modern Germanic languages); another group use them sparsely (in just a few functions and relatively infrequently, e.g., Mari or Hindi); finally another group, like Vietnamese and Tibetan, are highly co-compounding – they show a very high frequency of co-compounds in texts and use them in many different functions. All these parameters of variation are subject to Wälchli's investigation which involves several kinds of data, with the bulk of the data coming from texts – both original and parallel texts (the Universal Declaration of Human Rights and the Gospel according to Mark) in a large number of predominantly Eurasian languages. The data give rise to various generalizations on the patterning of co-compounds and their distribution across languages. Thus, for instance, it turns out that the meanings expressed by co-compounds in a language (its semantic profile) are intimately linked to their textual frequencies. On the other hand, the distribution of co-compounding across languages (including their semantic types and text frequencies) shows a macro-areal pattern of distribution, with a significant and steady decline from continental East and Southeast Asia westward in Eurasia.

Wälchli's study contains many interesting generalizations and explanations, a number of which are of primary methodological and theoretical relevance for future research, also on co-compounding (e.g., in non-Eurasian languages).

### 5.3 Lexical semantics in cross-linguistic research on motivation

As shown by the discussion of universal lexicalization in Section 4.6., there are very few meanings that can easily translate among languages, in particular if precise semantic identity is required. This fact is normally not explicitly considered in cross-linguistic research on motivation that usually takes meanings for granted, self-evident, and easily identified across languages and in a particular language. Consider the grammaticalization paths *HAND* → *FIVE*, attested in various languages, including Samoan (Polynesian, Austronesian) and Turkana (Nilotic, Nilo-Saharan), and *HAND* → *POSSESSION*, attested, among other languages, in Kono (Mande, Niger-Congo), Zande (Ubangian, Niger-Congo) (Heine & Kuteva 2002: 166–167) and Estonian (Finno-Ugric, Uralic). Since all these languages use the same lexeme for 'hand' and 'arm', how would we know that it is 'hand' that has been the grammaticalization source rather than, say, 'arm (excluding hand)' or 'arm (including hand)'? A strict proof for the case would include, first, arguments in favour of polysemy 'hand'/'arm' rather than semantic generality in all these languages and, second, evidence for the grammaticalized meanings being based on 'hand' to the exclusion of 'arm'. I am not aware of any serious attempts to do anything along these lines and doubt that there have been any. The interpretation of these particular examples and the postulation of these particular semantic links are, most probably, founded on common sense and intuition rather than on strict

argumentation, and on parallels with other languages which clearly distinguish between ‘hand’ and ‘arm’.

Likewise, almost none of the 18 verb pairs used in Nichols et al. (2004) and viewed by the authors as semantically basic and almost universals, belongs to Goddard’s (2001) list over lexico-semantic universals (with some, like ‘sit’ and ‘fear’, being explicitly excluded from it). The exact semantics and precise semantic identity of the verbs on the list is, however, not a point here: the 18 verb pairs have been chosen on pragmatic grounds, as representing certain combinations of general parameters, corresponding to frequently encoded situations and having approximate translational equivalents in many languages. Stricter requirements on semantic comparability would in fact create obstacles for achieving the principal objective of the study. Obtaining one-word expressions with the same semantics for 18 events (and, in addition, representing the various combinations of interesting parameters) in 80 languages is hardly conceivable, while choosing word combinations with the right semantics would most probably conceal the basic derivational relations.

In other cases, the relatively low degree of semantic precision in the definitions is less justified and can be impeding for deeper insights and effective cross-linguistic comparison. Among the various grammaticalization paths building on motion verbs several are often defined as starting with ‘come’ and ‘go’ (cf. in Heine & Kuteva 2002 for examples). The English verbs “come” and “go” as semantic metalabels are not totally felicitous; among other things, they encode the deictic distinction between centripetal and centrifugal motion, absent from many languages of the world (see Section 4.4. for the discussion of Ricca 1993) and neutralize the distinction between motion on foot vs. in a vehicle (cf. also Goddard 2001: 28). Descriptions like COME → CONTINUOUS, or GO → HABITUAL are therefore too vague for understanding the underlying logic of the development – they do, however, fulfill functions as preliminary crude classifications and as guidelines for future research.

The lack of consensus on the appropriate semantic meta-language and the form of meaning definitions creates obstacles for evaluating cross-linguistic connections even between studies of high semantic and lexicographic quality. Consider Enfield’s (2003) excellent book on the striking pattern of multi-functionality (polysemy and heterosemy) involving the verb ‘to acquire’ and shared by the languages of mainland Southeast Asia. The study suggests a fine-grained classification of the different meanings, illustrated by numerous relevant examples and provided with detailed semantic explications. Viberg (2002, 2006) also presents excellent studies on the semantically comparable verbs in European languages, primarily *få* in Swedish and *get* in English. In particular, *få* in Swedish shows an amazing diversity of uses, which has a clearly areal distribution (being replicated by its Norwegian cognate *få* and the etymologically unrelated verb *saada* in Finnish). Viberg (2006: 125) writes that “[e]ven if *få* has a relatively language-specific pattern of polysemy with respect to European languages in

general, it is not without parallels in other parts of the world” and mentions Enfield’s study. He concludes that “there is no exact parallel between the meanings of *få* and Lao *daj*<sup>4</sup> [“acquire”, MKT] but at a more general level the paths of extension are similar” (Viberg 2006: 126). I wish I could understand what this means. The very different ways of classifying phenomena and representing their meanings in Enfield’s and Viberg’s studies make it, in fact, very difficult to evaluate the degree of (dis)similarity between the two polysemy patterns (cf. Auwera et al. forthc. for a semantic map comparing parts of the relevant polysemy patterns – the North-European and Southeast Asian acquisitive modals, i.e., modals based on ‘get’/‘acquire’).

## 6. What cross-linguistic patterns are there in lexicon-grammar interaction?

Lehmann (1990: 163) defines lexical typology as research which focus on “typologically relevant features in the grammatical structure of the lexicon”, rather than on “the semantics of individual lexical items, their configurations in lexical field or individual processes of word formation” (Lehmann 1990: 165), i.e., the issues that have been considered as definitely belonging to lexical typological in the preceding sections.

Lehmann’s definition partly stems from a somewhat narrower understanding of typology than the one(s) suggested in Section 2. Typology, in this view, has to be based on essential properties that “vary regularly in the population under consideration” (Lehmann 1990: 165). As Lehmann explains this, the lexicon contains all that is completely idiosyncratic – and that does not therefore fit the premises for typological research – but also many regularities. It is a complex structure built upon categories and relations, with items falling into a number of lexical classes that are intimately connected to grammar. It is these typologically relevant features that are the primary object of the lexical typology. In a similar vein, Behrens & Sasse’s programmatic sketch (1997) promotes Lexical Typology (refers to a specific research framework), the aim of which is “to investigate cross-linguistically significant patterns of interaction between lexicon and grammar”, and mentions, among others, the following:

Viewed in the context of comparative linguistic research, the concept of lexicon-grammar leads to the assumption that we can expect, in different languages, quite divergent patterns of interaction between lexicon and grammar, and that these divergences are of great typological significance. It is therefore proposed that lexical semantics and its repercussions on grammar be assigned a central role in typological investigations. To this end, we will lay much emphasis on the discovery of principles of ambiguity and compositionality. These principles are presumably universal on a higher level of abstraction but typologically variable in their concrete individual manifestations. They therefore strongly influence the make-up of an individual language’s grammar and lexicon (Behrens & Sasse 1997: 1–2).

A number of various cross-linguistic studies can be attributed to lexical typology understood as a search for typologically relevant features in the grammatical structure of the lexicon, or as typologically significant correlations between lexicon and grammar. They vary in how and to what extent they fit into the typological research framework and tradition(s), and in how and to what extent they consider lexicon. On the whole, there is very little awareness that the relevant studies do focus on lexical phenomena. Some are restricted to lexicon-grammar interaction for a particular conceptual domain/lexical field or even for a particular lexical meaning, e.g., body-part terms in adnominal possession and in special syntactic constructions such as possessor ascension/external possession and body-part incorporation (Chappell & McGregor 1996; the literature is too extensive to be listed); kin terms in grammar (Dahl & Koptjevskaja-Tamm 2001); 'give' and argument linking (Haspelmath 2005a; Kittilä 2006), different classes of complement-taking verbs and the structure of complementation (Cristofaro 2003) 'want' and the structure of desiderative clauses (Haspelmath 2005b; Khanina 2005). Veselinova's (2006) large-scale study of suppletion in verb paradigms is an excellent example of lexicon-grammar interaction: it shows that suppletion tends to be linked to verbs with particular lexical meanings (e.g., MOTION), with different meanings picked up by suppletion according to different grammatical categories (e.g., tense-aspect-mood, or imperative). But many other traditional grammatical phenomena can be viewed as lexical.

First of all, consider the issue of word classes which has been subject to much debate and disagreements (with the relevant works being too numerous to be listed here). Word classes present an example par excellence of interaction – and significant correlation – between lexicon and grammar. The jump from individual language descriptions to large-scale cross-linguistic research tends, however, to reduce lexical information to very few representatives for each “presumptive” word class (like 'big' and 'good' for potential adjectives), not always systematically checked and/or completely comparable across the languages in the sample. Nonetheless, in a number of cross-linguistic works word-class behaviour is studied with more attention to lexical semantics and against the background of relatively fine-grained lexical distinctions. Thus, Dixon (1977) and later Dixon & Aikhenvald (2004) approach the problem of adjectives by breaking up the vague class of property concepts into a number of much more coherent semantic types (dimension, colour, physical properties, human propensities, etc.) which differ in their propensity to be lexicalized as “adjectives proper”, as verbs, as nouns or in still other ways. The suggested semantic classification and the labels for the semantic types are, in fact, less important than the much longer concrete lists of notions used more or less systematically across all the contributions in Dixon & Aikhenvald (2004). The specific language chapters provide various pieces of evidence that the notions included into one and the same semantic type vary in their propensity to be lexicalized as adjectives “proper” or in other ways, and a more fine-grained semantic classification might hopefully grasp the cross-linguistic systematicity here.

A radically “lexicon-based” stance is taken in Pustet's (2003) study of copulas in a global 131-language sample. Earlier work, primarily Stassen (1997), has suggested that copulas across languages show different inclination to combine with/to be required with different kinds of predicates along the hierarchy nominals > adjectivals > verbals. In other words, the first place where copulas occur in a language will be sentences like “Peter is a boy”, followed by “Peter is big”, with sentences like “Peter goes” having copulas rather infrequently. Pustet combines these generalizations with the various parameters suggested in earlier research as underlying the distinctions between verbs, adjectives and nouns (primarily in Croft 1991) and tests to what extent these are compatible. A part of her study is based on checking the behaviour of the items in large lexical samples (ranging from 530 to 850 items) as predicates in ten genetically, areally and structurally diverse languages. The lexical items fall into fourteen lexical classes based on various combinations of the three parameters of valence, transience and dynamicity which together define a three-dimensional semantic space and correlate with the presence vs. absence of copulas in a more principled and refined way than the earlier suggested hierarchy formulated in terms of word classes. It turns out that the majority of the lexical items in the sample represent just a small number of specific feature bundles – which, by and large, correspond to, or define the lexical prototypes of “entity”, or prototypical nominals (valence 0, –transient, –dynamic, e.g., 'house' and 'old man'), “property”, or prototypical adjectives (valence 1, ±transient, –dynamic, e.g., 'big'), and “event”, or prototypical intransitive and transitive verbs (valence 1 or 2, +transient, +dynamic, e.g., 'to go' and 'to buy'). The items within these each of these classes tend to show uniform behaviour with respect to copularization in a particular language. Lexical items from the other feature combinations, e.g., 'smart', 'hand' and 'son' (valence 1, –transient, –dynamic), 'to rain' (valence 0, +transient, +dynamic), 'to love' and 'to know' (valence 2, ±transient, –dynamic), share the behaviour of the “semantically adjacent” major classes, but only to a certain extent – the exact cuts-off between copularizing and non-copularizing lexemes are quite language-specific, even though governed by universal considerations (defined by the position of an item in the semantic space).

The fact that language-specific vocabularies include numerous minor (and normally small in size) classes, whose semantic profile does not coincide with those of prototypical nominals, adjectivals and verbals, is, thus, brought to the fore and taken seriously in Pustet's study. One of the theoretically important implications is also an identification of zones that show cross-linguistically recurrent overlaps in their word-class attribution, often even within one and the same language – e.g., terms denoting nationalities ('French'), emotional states ('to love'), bodily states ('to be tired'). Pustet's principled sample can be used for further work on word classes and on their interaction with various grammatical categories. And in general, future research on word classes will certainly benefit from taking lexical semantics seriously and working with more fine-grained lexical classes than what has often been done.

Cross-linguistic variation in categorization within major word classes also offers many opportunities for research on cross-linguistically significant patterns of interaction between lexicon and grammar. Verbs can be categorized in many different ways, for instance, depending on their argument structure, on how it is linked to a particular sentence structure and to what extent it can be subject to various valence alternations. The literature here is extensive and diverse, but is, in fact, very seldom explicitly linked to lexical phenomena, rather than being considered as primarily syntactic. As Nichols et al. (2004: 183) put it, “[i]f ergativity, for instance, were viewed as lexical, an ergative language would be one exhibiting ergativity as the default option or majority pattern in some broad category of verbs (e.g., all transitives, or some category of transitives), based on a standard sample of glosses. Stative-active alignment, viewed lexically, is lexically conditioned split intransitivity as Merlan (1985) presents it” (cf. Section 4.2. for the overview of Nichols et al.). A nice exception is Bossong’s (1998) study of experimenter constructions (like “I am cold”, “I am sorry”, “I see X”). It is built on the standard list of ten experience-denoting expressions across the European languages and shows significant genetic and areal differences in frequencies of the different syntactic patterns in which these expressions occur (among others, singling out the specific Standard Average European pattern).

A huge research domain of primary relevance for verbs focuses on the categories of aktionsart, aspect and tense. Although it is generally acknowledged that all these categories are very sensitive to the differences in the semantics of different verb classes, the valid cross-linguistic generalizations here are still quite few. Most modern research on aktionsart has its roots in Vendler’s (1967) verb classes (states, activities, accomplishments, achievements), whereas “[a]n urgent desideratum is the investigation of the role of lexicon, in particular the subcategorization of situation types”, as Sasse puts it in his overview of the recent development in the theory of aspect (2002: 263). Tatevosov’s (2002) study is a very promising step in this direction. It is based on a principled list of 100 “predicative meanings” (normally expressed by verbs or verb-based expressions) coming from several cognitive domains (from being and possession, motion, physical processes and changes, to phasal and modal verbs) and covering the “basic” verbal lexicon; these are checked for all possible combinations with the verbal tense-aspect categories and their resulting meanings in four genetically unrelated languages – Bagwalal (Daghestanian, NE Caucasian), Mari (Finno-Ugric, Uralic), Tatar (Turkic) and Russian (Slavic, Indo-European). Already this comparison falsifies the common assumptions “that notions on which Vendlerian classes are based are logically universal, hence are not subject to crosslinguistic variation” and that verbs or verb phrases with “similar meanings” in different languages (i.e., translational equivalents) will belong to the same verbal class as their English equivalents (Tatevosov’s 2002: 322).

“Actionality”, used by Tatevosov instead of “Aktionsart”, turns thus out to be a parameter based on a universal set of elementary semantic distinctions, but allowing

for different settings in different languages. Different languages show therefore their own language specific subcategorizations of the verb lexicon that can only be discovered via empirical investigations rather than taken for granted. A particularly beautiful example of the latter comes from Botne’s (2003) cross-linguistic study of *die* and its correspondences in 18 languages. ‘Die’ is always quoted as the prototypical example of Vendler’s achievement verbs (telic, or bounded, and punctual) in that it refers to the acute point demarcating life and death. Botne shows, however, that languages can differ in their lexicalization of the different stages in the process leading to death, which, in turn, has important consequences for the aktionsart categorization of the corresponding verb in a particular language.

Cross-linguistic variation in categorization within nouns also offers many interesting topics for research on lexicon-grammar interaction. For instance, subcategorization of nouns according to their interaction with the category of number, involves various fascinating and cross-linguistically still poorly understood issues such as count-mass distinction, collectives, singular vs. pluralia tantum, etc. (for some examples see Corbett (2000); Wierzbicka (1988); Rijkhoff’s (2002) notion of “Seinsarten”, Koptjevskaja-Tamm & Wälchli’s (2001) areal-typological study of pluralia tantum in the Circum-Baltic area against a broader European context, Lucy (1992), Behrens (1995) and Koptjevskaja-Tamm (2004) on count-mass distinctions across languages). Table 6 gives a flavour of how differently count-mass categorization can work in relatively closely related languages (and, in addition, in cognates).

**Table 6.** Some examples of count-mass differences in Russian, German, Swedish and Italian

Count (+) or mass (-)	Russian	German	Swedish	Italian
strawberry	- <i>klubnika</i>	+ <i>Erdbeer</i>	+ <i>jordgubbe</i>	+ <i>fragola</i>
fruit	+ <i>frukt</i>	- <i>Obst</i>	± <i>frukt</i>	+ <i>frutto</i>
hair	+ <i>volos</i>	± <i>Haar</i>	- <i>hår</i>	+ <i>capello</i>
furniture	- <i>mebel'</i>	+ <i>Möbel</i>	+ <i>möbel</i>	+ <i>mobile</i>

Possible implications of such variation for the lexical semantics of the items under consideration are very rarely explicitly acknowledged and discussed in cross-linguistic studies on lexicon-grammar interaction. Consider Botne’s (2003) conclusions following his cross-linguistic study on the aktionsart categorization of the correspondences to *die*:

This small, exploratory study has shown that ... achievement verbs, though unified by the punctual, culminative nature of their nucleus, may be conceptualized in

different languages as encoding durative preliminary (onset) or postliminary (coda) phases in addition to the punctual nucleus. Consequently, die verbs frequently have a complex temporal structure and do not simply encode a point of transition ... [T]he same “concept” will not necessarily be encoded with the same phases in every language. Consequently, appropriate cross-linguistic comparison and analysis of these kinds of verbs will perforce require a close analysis of a particular verb in each language (Botne 2003: 276).

But if languages differ as to which of the phases leading to death they encode in their ‘die’-verbs, can we still view them as encoding “the same concept”? Likewise, the meaning of the German mass noun *Obst* is hardly identical to that of the Russian count noun *frukt*, even though they constitute translational equivalents to each other. Cross-linguistic identification of phenomena based on “approximate”, rather than “precise” semantic identity, can be justified when the primary focus of the cross-linguistic research is not on the lexical semantics per se (cf. with the discussion in Section 5.3.). However, it is also reasonable to take the next step and use the cross-linguistic variation in grammatical behaviour as evidence for the lexical-semantic differences.

It is now widely acknowledged by various linguistic theories that a large portion of grammatical phenomena is rooted in the lexicon. Lexicon-grammar interaction will surely provide lots of challenges for the future lexical-typological research.

## 7. Lexical typology: Past, present and future

It is impossible to cover all the aspects of lexical-typological research in one paper. One important group of questions that have not been touched upon concerns cross-linguistically recurrent patterns in contact-induced lexicalization and lexical change: e.g., differences in borrowability among the different parts of the lexicon and the corresponding processes in the integration of new words, or patterns of lexical acculturation (i.e., how lexica adjust to new objects and concepts). The important contributions here include Brown (1999) and the on-going project on “Loanword Typology: toward the comparative study of lexical borrowability in the world’s languages” at Max Planck Institute for Evolutionary Anthropology (coordinated by Martin Haspelmath and Uri Tadmor, <http://www.eva.mpg.de/lingua/files/lwt.html>), which modify various traditional assumptions (like non-borrowability, or at least, “relative” non-borrowability of the items on Swadesh’ “basic vocabulary” lists), come up with new cross-linguistic generalizations and suggest methodology for future research.

Issues related to the “basic vocabulary” (in various understandings) and the “lexico-typological profile” of a language figure also in other connections in cross-linguistic research with lexical-typological ambitions (Viberg 2006; Koch & Marzo 2007; Kibrik 2003). There are also interesting questions on the interaction between lexicon and

phonology, overall principles of taxonomic categorization or the organization of the lexicon, and surely many others.

Possible theoretical implications of lexical typology are vast – for theoretical linguistics in general (cf. Koptjevskaja-Tamm et al. 2007 for some details) and for broader issues such as child language acquisition (with Melissa Bowerman and Dan Slobin as the leading authorities) and “Linguistic Relativity”.

In other words, there are numerous diverse and fascinating questions for the future lexical-typological research. Its basic and most urgent problems are primarily methodological. For instance, we need to:

- refine the existent methods of data collection and develop new ones, improve standards in cross-linguistic identification of studied phenomena and in their (semantic) analysis,
- achieve a reasonable consensus on the meta-language used for semantic explications and on the ways of representing meanings.

To start with the first issue, the methodology of data collection. Morphosyntactic typology has been largely dependent on secondary data sources, with reference grammars as the undoubtedly most often used data source, in many cases complemented by sporadic consultations with native speakers and/or language experts. Studies in morphosyntactic typology are typically a “one researcher’s job”: even when data collection involves filling in questionnaires and responding to other data elicitation stimuli, the people doing that part of job normally count as consultants, rather than co-authors (some of the exceptions being the tradition of the Leningrad/St.Peterburg Typological School, or the numerous collections edited by Aikhenvald and Dixon).

For lexical typology, on the contrary, secondary sources are of marginal importance, in particular, if we take the three groups of questions that have been the main subject matter of this paper – categorization within conceptual domains, semantic and formal motivation, lexicon-grammar interaction. Relevant data are normally scattered across different kinds of secondary sources: a thesaurus might provide information on categorization within conceptual domains, while a “normal” dictionary may have something on the polysemy patterns and other formal-semantic relations within word families. Some information on lexicon-grammar interaction might be occasionally given in a reference grammar (which seldom lists all the words showing a particular grammatical behaviour), some might be appear in a dictionary. A desideratum would be to have a source that for every word in a language would give a precise meaning definition, show both its exact relations to other words and define its grammatical properties. There are a few attempts on the market towards this desideratum for the better described languages – e.g., “The interpretational-combinatorial dictionary” with roots in the Moscow School of Semantics (cf. Iordanskaja & Paperno 1996 for the excellent treatment of the Russian body-part terms in this tradition), the Berkeley

FrameNet project based on Frame Semantics ([http://framenet.icsi.berkeley.edu/index.php?option=com\\_frontpage&Itemid=1](http://framenet.icsi.berkeley.edu/index.php?option=com_frontpage&Itemid=1))<sup>3</sup> or the expanding enterprise of WordNet for several European languages (<http://www.globalwordnet.org/>).<sup>4</sup> However, the lexicon for most languages of the world is – and will remain – relatively poorly described, at least for the purposes of consistent cross-linguistic research.

Most lexical-typological research is therefore in need of constantly inventing, testing and elaborating its methods of data collection. Even more, the different methods are not easily transmittable among different research areas, even those that ask comparable questions. For instance, visual stimuli for eliciting words referring to CUTTING AND BREAKING EVENTS can certainly serve as a model for research on some other conceptual domains involving dynamic situations with clearly visible actions and results (say, DRESSING / UNDESSING, or PUTTING). But already moving to domains based on other perceptual modalities is far from trivial – SOUNDS, TEMPERATURE, TASTE are still awaiting good data collection techniques and guidelines – whereas the EMOTIONAL and MENTAL world is even harder to cover with perceptual stimuli (cf., however, Pavlenko 2002 for a comparison of emotional descriptions in Russian and English narratives elicited through the same short film).

The extent to which parallel texts can be used in lexical-typological research is, of course, extremely dependent on the object of study and on the genre of the parallel texts and is best suited for frequent phenomena. Thus, while motion verbs frequently occur in the New Testament, and generic statements in the Universal Declaration of the Human Rights, which are the easiest available texts in many languages, these sources will be of restricted value for the study of PAIN expressions, even though such examples do occasionally occur in the former. However, for a more limited number of languages other parallel texts have been successfully used in lexical-typological work (“Le petit prince” and “Harry Potter” belong to the favourites here).

Finally, word lists, as we have seen, may well be used for some purposes (e.g., for checking the word-class categorization of “property” words or the aktionsart categorization of verbs, etc.), but are of marginal value when too little is known about the lexical meaning of phenomena under consideration or when the phenomena involve too many language-specific lexical idiosyncrasies. Consider *pluralia tantum*, e.g., nouns that only occur in the plural form, like *scissors*, and are very unevenly distributed across languages. In Koptjevskaja-Tamm & Wälchli (2001) we used two principled samples of lexemes that are encoded by pluralia-tantum nouns in Lithuanian vs. Russian for collecting comparable data across forty European languages. Since the distribution of pluralia tantum in a language is highly idiosyncratic, we hypothesized that

3. Last visited on April 26th 2007.

4. Last visited on April 26th 2007.

the degree of overlapping in the distribution of pluralia tantum across languages could be used as a measure for their contacts, in this case, the languages in the Northeastern part of Europe. While the two samples turned out to be useful for this particular end, the same fact (lexical idiosyncrasies of pluralia tantum) causes difficulties for cross-linguistic studies of pluralia tantum in general. Although they do often occur in comparable domains (e.g., heterogeneous substances, like *leftovers*, diseases, like *measles*, festivities, like *Weihnachten* “Christmas” in German), they are very language-specific when it comes to the lexical meanings, which rules out the use of a consistent word list for cross-linguistic data collection.

In my opinion, successful lexical-typological research should in most cases build on a collaborative work involving language experts (and, possibly, other specialists as well). The work on lexical universals within the NSM tradition (Goddard & Wierzbicka 1994), the different domain-categorization studies co-ordinated from the Max-Planck Institute in Nijmegen (Levinson & Meira 2003; Majid et al. eds. 2006; Majid & Bowerman eds. 2007), the project on aquamotion verbs directed by Moscow linguists (Maisak & Rakhilina 2007) are all examples of excellent semantic-typological research based on the methodology that had been elaborated, tested and improved by the group of language experts, who have further collected and analyzed the data in close collaboration with native speakers.

The issue of data collection is, of course, intimately related to the issue of cross-linguistic identification of studied phenomena, which is a key concern for cross-linguistic and typological research in general. We have to be sure that we compare like with like, rather than apples with pears. However, another key concern for cross-linguistic and typological research is to find a reasonable level of abstraction, at which the richness of language-specific details can be reduced to manageable patterns. The two concerns interact in various ways; most importantly, what counts as “like and like” is often dependent on the research object and goal. In the course of this paper we have had several occasions to discuss the different levels of semantic precision appropriate for different types of lexical-typological research (e.g., domain-categorization vs. semantic motivation vs. lexicon-grammar interaction). It should be mentioned here that the grammatical typology on the whole hardly ever cares about precise semantics: the only prerequisite is that we can roughly identify linguistic phenomena across languages via certain conditions that they have to meet, e.g., via a certain function that has to be expressed by a construction. Thus, for instance, an possessive NP is recognized by its ability to refer to legal ownership (*Peter’s bag*), to kin relations (*Peter’s son*) or to relations between a person and his body-parts (*Peter’s leg*). The fact that the same construction in English can occasionally refer to temporal and local relations (*yesterday’s magazine*, *London’s museums*), whereas many other languages are much more restrictive in this respect is of marginal interest for the cross-linguistic identification of possessive NPs themselves. There are, of course, certain limits to the semantic

vagueness that can underlie systematic cross-linguistic identification of phenomena. I find it difficult to set up good methods for testing universality of some suggested metaphors cross-linguistically, like, for instance, ANGER IS HEAT (Kövecses 1996). What can probably be done is to test some of its specific manifestations, e.g., whether the words for anger (and other emotions) can be described by temperature terms.

Finally, in order to achieve a reasonable consensus on the meta-language used for semantic explications and on the ways of representing meanings is an urgent need – both in theoretical semantics, in semantic and lexical typology and in lexicography.

Let's hope that the contributions in this volume will be good points of departure for numerous future projects in lexical typology.

### Acknowledgements

This paper has a long prehistory. Parts of it have been presented in different versions at the Annual Meeting of the Finnish and Estonian Linguistic Society in Tallinn (May 2004), at the conference “Lexicon in linguistic theory” (Åbo/Turku, November 2004), at the workshop “Semantic parallels in a cross-linguistic perspective” (Paris, December 2004), in lectures at the universities of Stockholm, Gothenburg, Lund, Pavia, Turin and Tübingen. I am grateful to the audiences for their questions and comments. I would also like to thank several friends and colleagues who have provided intellectual and humane support, inspiration and feedback in my work on the paper: Melissa Bowerman, Grev Corbett, Östen Dahl, Nick Enfield, Nick Evans, Giannoula Giannouloupoulou, Cliff Goddard, Olesya Khanina, Seppo Kittilä, Peter Koch, Christian Lehmann, Eva Lindström, Galina Paramei, Ekaterina Rakhilina, Edith Moravcsik, Carita Paradis, Frans Plank, Farzad Sharifian, Jae Jung Song, Martin Tamm, Martine Vanhove, Ljuba Veselinova, Bernhard Wälchli. All the faults remain mine. There is a certain overlapping between this paper and Koptjevskaja-Tamm et al. (2007). This concerns primarily parts of Section 6 and Section 7.

### References

- Aikhenvald, A. & Dixon, R.M.W. (Eds). 2002. *Word: A Cross-linguistic Typology*. Cambridge: CUP.  
 Ameka, F. & Levison, S. (Eds). 2007. Locative predicates. *Linguistics*, special issue 45(5–6).  
 Andersen, E. 1978. Lexical universals of body-part terminology. In *Universals of Human Language*, J. Greenberg (Ed.), 335–368. Stanford CAC Stanford University Press.  
 Auwera, J. van der. 1998. Phrasal adverbials in the languages of Europe. In *Adverbial Constructions in the Languages of Europe*, J. van der Auwera (Ed.), 25–145. Berlin: Mouton de Gruyter.  
 Auwera, J. van der. 2001. On the lexical typology of modals, quantifiers, and connectives. In *Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Kiefer*, I. Kenesei & R.M. Harnish (Eds), 173–186. Amsterdam: John Benjamins.

- Auwera, J., van der, Kehayov, P. & Vittrant, A. Forthcoming. Modality's semantic map revisited: Acquisitive modals. In *Cross-linguistic Studies of Tense, Aspect, and Modality*, L. Hogeweg, H. De Hoop & A. Malchukov (Eds). Amsterdam: John Benjamins.  
 Bach, E., Jelinek, E., Kratzer, A. & Partee, B.H. (Eds). 1995. *Quantification in Natural Languages*. Dordrecht: Kluwer.  
 Behrens, L. 1995. Categorizing between lexicon and grammar. The Mass/Count distinction in a cross-linguistic perspective. *Lexicology* 1(1): 1–112.  
 Behrens, L. & Sasse, H.J. 1997. *Lexical Typology: A Programmatic Sketch* [Arbeitspapier Nr. 30, Neue Folge]. Cologne: University of Cologne, Department of linguistics.  
 Berlin, B. 1992. *Ethnobiological Classification*. Princeton NJ: Princeton University Press.  
 Berlin, B. & Kay, P. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley CA: University of California Press.  
 Berman, R.A. & Slobin, D.I., in collaboration with Aksu-Koç, A. 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale NJ: Lawrence Erlbaum Associates.  
 Blank, A. 2001. Pathways of lexicalization. In Haspelmath et al. (Eds), 1596–1608.  
 Bossong, G. 1998. Le marquage de l'expérience dans les langues d'Europe. In *Actance et valence dans les langues de l'Europe*, J. Feuillet (Ed.), 259–294. Berlin: Mouton de Gruyter.  
 Botne, R. 2003. To die across languages. *Linguistic Typology* 7(2): 233–278.  
 Bowden J. 1992. *Behind the Preposition: the grammaticalisation of locatives in Oceanic languages*. Canberra: Pacific Linguistics.  
 Brown, C.H. 1976. General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist* 3: 400–424.  
 Brown, C.H. 1999. *Lexical Acculturation in Native American Languages*. Oxford: OUP.  
 Brown, C.H. 2001. Lexical typology from an anthropological point of view. In Haspelmath et al. (Eds), 1178–1190.  
 Brown, C.H. 2005a. Hand and arm. In Haspelmath et al. (Eds), 522–525.  
 Brown, C.H. 2005b. Finger and hand. In Haspelmath et al. (Eds), 526–529.  
 Chappell, H. & McGregor, W. (Eds). 1996. *The Grammar of Inalienability. A Typological Perspective on Body Part Terms and the Part-whole Relation*. Berlin: Mouton de Gruyter.  
 Corbett, G.G. 2007. Canonical typology, suppletion and possible words. *Language* 83(1): 8–42.  
 Corbett G.G. 2000. *Number* [Cambridge Textbooks in Linguistics]. Cambridge: CUP.  
 Corbett G.G. Forthcoming. Higher order exceptionality in inflectional morphology. In *Expecting the Unexpected: Exceptions in Grammar*, H.J. Simon & H. Wiese (Eds). Berlin: Mouton de Gruyter.  
 Cristofaro, S. 2003. *Complementation*. Oxford: OUP.  
 Croft, W. 1990[2003]. *Typology and Universals* [Cambridge Textbooks in Linguistics], 1<sup>st</sup>-2<sup>nd</sup> Edn. Cambridge: CUP.  
 Croft, W. 1991. *Syntactic Categories and Grammatical Relations*. Chicago IL: Chicago University Press.  
 Cruse, A. 1986. *Lexical Semantics*. Cambridge: CUP.  
 Cruse, A., Hundsnerscher, F., Job, M. & Lutzeier, P.R. (Eds). 2002[2005]. *Lexicology. An International Handbook on the Nature and Structure of Words and Vocabularies*, Vols. 1–2. Berlin: Walter de Gruyter.  
 Cysouw, M. 2004. Interrogative words: An exercise in lexical typology. Ms, (<http://email.eva.mpg.de/~cysouw/presentations.html>)  
 Dahl, Ö. & Koptjevskaja-Tamm, M. 2001. Kinship in grammar. In *Dimensions of Possession*, I. Baron, M. Herslund & F. Sorensen (Eds), 201–225. Amsterdam: John Benjamins.

- Dedrick, D. 2005. Color, color terms, categorization, cognition, culture: An afterthought. *Journal of cognition and culture* 5(3–4): 487–495.
- Dixon, R.M.W. 1977. Where have all the adjectives gone? *Studies in Language* 1: 19–80.
- Dixon, R.M.W. & Aikhenvald, A. (Eds). 2004. *Adjective Classes*. Oxford: OUP.
- Enfield, N.J. 2003. *Linguistic epidemiology*. London & New York: Routledge/Curzon.
- Enfield, N.J. 2002. Semantic analysis of body parts in emotion terminology: Avoiding the exoticisms of “obstinate monosemy” and “online extension”. *Pragmatics and Cognition* 10(1–2): 81–102.
- Enfield, N.J. 2007. Lao separation verbs and the logic of linguistic event categorization. In Majid & Bowerman (Eds), 287–286.
- Enfield, N.J., Majid, A. & van Staden, M. 2006. Cross-linguistic categorisation of the body: Introduction. In Majid et al. (Eds), 134–147.
- Enfield, N.J. & Wierzbicka, A. (Eds). 2002. The body in description of emotion. *Pragmatics and cognition* 10(1–2) (special issue).
- Evans, N. 2001. Kinship terminology. In *International Encyclopedia of Social and Behavioral Sciences*, N.J. Smelser & P. Baltes (Eds), 8105–8111. Amsterdam: Elsevier Science.
- Evans, N. 1992. Multiple semiotic systems, hyperpolysemy, and the reconstruction of semantic change in Australian languages. In *Diachrony within Synchrony*, G. Kellerman & M. Morrissey, (Eds), 475–508. Bern: Peter Lang.
- Evans, N. Forthcoming. Semantic typology. In *The Oxford Handbook of Typology*, J.J. Sung (Ed.), Oxford: OUP.
- Evans, N. & Sasse, H.J. 2007. Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. *Archives and Social Studies: A Journal of Interdisciplinary Research* 1. ([http://socialstudies.cartagena.es/index.php?option=com\\_content&task=view&id=53&Itemid=42](http://socialstudies.cartagena.es/index.php?option=com_content&task=view&id=53&Itemid=42))
- Evans, N. & Wilkins, D.P. 2000. In the mind’s ear: The semantic extensions of perception verbs in Australian languages. *Language* 76: 546–592.
- Foley, W. 1997. *Anthropological Linguistics. An Introduction*. Malden MA: Blackwell.
- Gil, D. 2001. Quantifiers. In Haspelmath et al. (Eds), Vol. 2, 1275–1294.
- Goddard, C. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5: 1–65.
- Goddard, C. & Wierzbicka, A. (Eds). 1994. *Semantic and Lexical Universals*. Amsterdam: John Benjamins.
- Grandi, N. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: F. Angeli.
- Greenberg, J.H. 1966. *Language Universals* [Janua Linguarum, Series Minor 59]. The Hague: Mouton.
- Greenberg, J.H. 1980. Universals of kinship terminology. In *On Linguistic Anthropology: Essays in Honor of Harry Hoiyer*, J. Maquet (Ed.), 9–32. Malibu CA: Udena Publications.
- Harkins, J. & Wierzbicka, A. (Eds). 2001. *Emotions in Cross-linguistic Perspective*. Berlin: Mouton de Gruyter.
- Haspelmath, M. 1987. *Transitivity Alternations of the Anticausative Type* [Arbeitspapier Nr. 5, Neue Folge]. Cologne: University of Cologne.
- Haspelmath, M. 1993. Inchoative/causative verb alternations. In *Causatives and Transitivity*, B. Comrie & M. Polinsky (Eds), 87–111. Amsterdam: John Benjamins.
- Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: OUP.
- Haspelmath, M. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In *The New Psychology of Language*, M. Tomasello (Ed.), Vol. 2, 211–242. Mahwah NJ: Lawrence Erlbaum Associates.
- Haspelmath, M. 2005a. Ditransitive constructions: The verb ‘give’. In Haspelmath et al., 426–429.
- Haspelmath, M. 2005b. ‘Want’ complement clauses. In Haspelmath et al., 502–505.
- Haspelmath, M., Dryer, M., Gil, D. & Comrie, B. 2005. *The World Atlas of Language Structures (WALS)*. Oxford: OUP.
- Haspelmath, M., König, E., Oesterreicher, W. & Raible, W. (Eds). 2001. *Language Typology and Language Universals*, Vol.1–2. Berlin: Walter de Gruyter.
- Haspelmath, M. & Tadmor, U. (Eds) In preparation. *Lexical Borrowability*. 2 Vols. Berlin: Mouton de Gruyter.
- Heine, B. 1989. Adpositions in African languages. *Linguistique Africaine* 2: 77–127.
- Heine, B. 1997. *Cognitive Foundations of Grammar*. Oxford: OUP.
- Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP.
- Iordanskaja, L. & Paperno, S. 1996. *A Russian-English Collocational Dictionary of the Human Body*. Columbus OH: Slavica Publishers. (Also <http://LexiconBrige.Com/for> the electronic edition)
- Jameson, K. 2005. Culture and cognition: What is universal about the representation of color experience? *Journal of Cognition and Culture* 5(3–4): 294–347.
- Juravsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3): 533–578.
- Kay, P. & Maffi, L. 2000. Color appearance and the emergence and evolution of basic color naming. *Journal of linguistic anthropology* 1: 12–25.
- Kay, P. & Maffi, L. 2005. Colour terms. In Haspelmath et al. (Eds), 534–545.
- Kay, P. & McDaniel, C. 1978. The linguistic significance of the meanings of basic color terms. *Language* 54: 610–46.
- Kemmer, S. 1993. *The Middle Voice*. Amsterdam: John Benjamins.
- Khanina, O. 2005. *Jazykove oformlenie situacii zelanija: opyt tipologičeskogo issledovanija* [Linguistic Encoding of ‘Wanting’: A Typological Approach]. Ph.D. dissertation, Moscow State University.
- Khanina, O. Forthcoming. How universal is wanting? *Studies in Language*.
- Kibrik, A.A. 2003. Lexical semantics as a key to typological peculiarities. Paper at the Workshop on Lexical/Semantic Typology at the ALT conference in Cagliari, September 2003.
- Kita, S. & Özyürek, A. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48: 16–32.
- Kittilä, S. 2006. Anomaly of the verb ‘give’ explained by its high (formal and semantic) transitivity. *Linguistics* 44(3): 569–609.
- Koch, P. 2001. Lexical typology from a cognitive and linguistic point of view. In Haspelmath et al. (Eds), Vol. 2, 1142–1178.
- Koch, P. & Marzo, D. 2007. A two-dimensional approach to the study of motivation and lexical typology and its first application to French high-frequency vocabulary. *Studies in Language* 31(2): 259–291.
- Koptjevskaja-Tamm, M. 2004. Mass and collection. In *Morphology: A Handbook on Inflection and Word Formation*, G. Booij, C. Lehmann & J. Mugdan (Eds), Vol. 2, 1016–1031. Berlin: Walter de Gruyter.
- Koptjevskaja-Tamm, M., Divjak, D. & Rakhilina, E. Forthcoming. Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In *Multiple Perspectives on Slavic Verbs of Motion*, V. Dragina & R. Perelmutter (Eds). Amsterdam: John Benjamins.
- Koptjevskaja-Tamm, M. & Rakhilina, E. 2006. “Some like it hot”: On semantics of temperature adjectives in Russian and Swedish. *STUF (Sprachtypologie und Universalienforschung)*, a

- special issue on *Lexicon in a Typological and Contrastive Perspective*, G. Giannouloupoulou & T. Leuschner (Eds) 59(2): 253–269.
- Koptjevskaja-Tamm, M., Vanhove, M. & Koch, P. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11(1): 159–186.
- Koptjevskaja-Tamm, M. & Wälchli, B. 2001. The Circum-Baltic languages: An areal-typological approach. In *Circum-Baltic Languages: Their Typology and Contacts*, Ö. Dahl & M. Koptjevskaja-Tamm (Eds), Vol. 2, 615–750. Amsterdam: John Benjamins.
- Krifka, M. (Ed.). 2003. *Natural Semantic Metalanguage*. A special issue of *Theoretical Linguistics* 29(3). Berlin: Mouton de Gruyter.
- Kövecses, Z. 1996. Anger: Its language, conceptualization, and physiology in the light of cross-cultural evidence. In *Language and the Cognitive Construal of the World*, J. Taylor & R. MacLaury (Eds), 181–196. Berlin: Mouton de Gruyter.
- Lang, E. 2001. Spatial dimension terms. In Haspelmath et al. (Eds), 1251–1275.
- Lehmann, C. 1990. Towards lexical typology. In *Studies in Typology and Diachrony. Papers Presented to Joseph H. Greenberg on his 75th Birthday*, W. Croft, K. Denning & S. Kemmer (Eds), 161–185. Amsterdam: John Benjamins.
- Lehrer, A. 1992. A theory of vocabulary structure: Retrospectives and prospectives. In *Thirty Years of Linguistic Evolution. Studies in Honour of René Dirven on the Occasion of his Sixtieth Birthday*, M. Pütz (Ed.), 243–256. Amsterdam: John Benjamins.
- Levinson, S. 2006. Parts of the body in Yéli Dnye, the Papuan language of Rossel Island. In *Parts of the Body: Cross-linguistic Categorization*, N. Enfield, A. Majid & M. van Staden (Eds), special issue of *Language Sciences* 28: 221–240.
- Levinson, S. & Meira, S. 2003. “Natural concepts” in the spatial topological domain – adpositional meanings in cross-linguistic perspective: An exercise in semantic typology. *Language* 79(3): 485–516.
- Lichtenberk, F. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67(3): 475–509.
- Lindström, E. 2002. The body in expressions of emotion: Kuot. In Enfield, & Wierzbicka (Eds), 159–184.
- Lucy, J.A. 1992. *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Cambridge: CUP.
- MacLaury, R.E. 1997. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Austin TX: University of Texas Press.
- MacLaury, R.E. 2001. Color terms. In Haspelmath et al. (Eds), 1227–1251.
- Maisak, T. & Rakhilina, E. (Eds). 2007. *Glagoly dviženija v vode: leksičeskaja tipologija* (Aquamation Verbs: Lexical Typology). Moskva: Indrik.
- Majid, A., Enfield, N.J., & van Staden, M. (Eds). 2006. Parts of the body: Cross-linguistic categorization. *Language Sciences* 28(2–3): 137–360 (special issue).
- Majid, A. & Bowerman, M. (Eds). 2007. Cutting and breaking events: A cross-linguistic perspective. *Cognitive Linguistics* 18(2): 153–177 (special issue).
- Majid, A., Bowerman, M., van Staden, M. & Boster, J.S. (Eds). 2007. The semantic categories of ‘cutting and breaking’ events across languages. *Cognitive Linguistics* 18(2): 133–152.
- Maslova, E. 2004. A universal constraint on the sensory lexicon, or when hear can mean ‘see’? In *Tipologičeskie obosnovanija v grammatike: k 70-letiju professora Xrakovskogo V.S.* (Typological Foundations in Grammar: For Professor Xrakovskij on the Occasion of his 70<sup>th</sup> Birthday), A.P. Volodin (Ed.), 300–312. Moscow: Znak. (Also <http://www.stanford.edu/~emaslova/Publications/PubFrame.html>)
- Matisoff, J. 1986. Hearts and minds in Southeast Asian languages and English: An essay in the comparative lexical semantics of psycho-collocations. *Cahiers de linguistique Asie Orientale* 15(1): 5–57.
- Merlan, F. 1985. Split intransitivity: Functional oppositions in intransitive inflection. In *Grammar Inside and Outside the Clause: Some Approaches to the Theory from the Field*, J. Nichols & A. Woodbury (Eds), 324–362. Cambridge: CUP.
- Morgan, L.H. 1870. *Systems of Consanguinity and Affinity*. (Reprinted in 1997, Lincoln NB: University of Nebraska Press).
- Nerlove, S. & Romney, A.K. 1967. Sibling terminology and cross-sex behavior. *American Anthropologist* 74: 1249–1253.
- Newman J. 2002. *The Linguistics of Sitting, Standing and Lying*. Amsterdam: John Benjamins.
- Nichols, J., Peterson, D.A. & Barnes, J. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8: 149–211.
- Ojutkangas, K. 2000. Grammatical possessive constructions in Finnic: *Käsi* ‘hand’ in Estonian and Finnish. In *Facing Finnic. Some Challenges to Historical and Contact Linguistics* [Castrenianumin toimitteita 59], J. Laakso (Ed.), 137–155. Helsinki: Finno-Ugrian Society – Dept. of Finno-Ugrian Studies of the University of Helsinki.
- Pavlenko, G. 2002. Emotions and the body in Russian and English. In Enfield & Wierzbicka (Eds), 207–241.
- Paramei, G. 2005. Singing the Russian blues: An argument for culturally basic color terms. *Cross-cultural research* 39(1): 10–34.
- Payne, D. 2006. Color terms. In *Encyclopedia of Languages and Linguistics*, 601–605. Oxford: Elsevier.
- Pustet, R. 2003. *Copulas. Universals in the Categorization of the Lexicon*. Oxford: OUP.
- Quine, W.V. 1960. *Word and Object*. Cambridge MA: The MIT Press.
- Rakhilina, E. 2006. *Glagoly plavanija v russkom jazyke* (Aquamation verbs in Russian). In Maisak & Rakhilina (Eds), 267–285.
- Riemer, N. 2005. *The Semantics of Polysemy*. Berlin: Mouton de Gruyter.
- Ricca, D. 1993. *I verbi deittici di movimento in Europa: Una ricerca interlinguistica*. Firenze: La Nuova Italia Editrice.
- Rijkhoff, J. 2002. *The Noun Phrase* [Oxford Studies in Typology and Linguistic Theory] Oxford: OUP.
- Sasse, H.J. 2002. Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology* 6(2): 199–271.
- Schladt M. 2000. The typology and grammaticalization of reflexives. In *Reflexives: Forms and Functions*, Z. Frajzyngier (Ed.), 103–124. Amsterdam: John Benjamins.
- Sharifian, F., Dirven, R., Yu, N. & Neiemier, S. (Eds). Forthcoming. *Culture, Body, and Language: Conceptualizations of Internal Body Organs Across Cultures and Languages*. Berlin: Mouton De Gruyter.
- Slobin, D.I. 2003. Language and thought online: Cognitive consequences of linguistic relativity. In *Language in Mind: Advances in the Study of Language and Thought*, D. Gentner & S. Goldin-Meadow (Eds), 157–191. Cambridge MA: The MIT Press.
- Slobin, D.I. & Bowerman, M. 2007. Interfaces between linguistic typology and child language research. *Linguistic Typology* 11: 213–226.
- Slobin, D.I. & Hoiting, N. 1994. Reference to movement in spoken and signed languages: Typological considerations. *BLS* 20: 487–505.
- Stassen, L. 1997. *Intransitive Predication*. Oxford: OUP.

- Svorou, S. 1993. *The Grammar of Space*. Amsterdam: John Benjamins.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: CUP.
- Talmy, L. 1991. Path to realization – Via aspect and result. A typology of event conflation. *BLS* 17: 480–519.
- Talmy, L. 1985. Lexicalization patterns. In *Language Typology and Synchronic Description*, T. Shopen (Ed.), Vol. 3, 57–149. Cambridge: CUP.
- Tatevosov, S. 2002. The parameter of actionality. *Linguistic Typology* 6(3): 317–401.
- Terrill, A. 2006. Body-part terms in Lavukaleve, a Papua language of the Solomon Islands. In Majid et al. (Eds), 304–322.
- Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca NY: Cornell University Press.
- Veselinova, L. 2006. *Suppletion in Verb Paradigms*. Amsterdam: John Benjamins.
- Viberg, Å. 1984. The verbs of perception: a typological study. *Linguistics* 21: 123–162.
- Viberg, Å. 2001. Verbs of perception. In Haspelmath et al. (Eds), Vol. 2, 1294–1309.
- Viberg, Å. 2002. Polysemy and disambiguation cues across languages. The case of Swedish *få* and English *get*. In *Lexis in Contrast*, B. Altenberg & S. Granger (Eds), 119–150. Amsterdam: John Benjamins.
- Viberg, Å. 2005. The lexical typological profile of Swedish mental verbs. *Languages in contrast* 5(1): 121–157.
- Viberg, Å. 2006. Towards a lexical profile of the Swedish verb lexicon. *Sprachtypologie und Universalienforschung (STUF)* 59(1): 103–129. (special issue on *The Typological Profile of Swedish*, Å. Viberg (Ed.)).
- Wälchli, B. 2005. *Co-compounds and Natural Coordination*. Oxford: OUP.
- Wälchli, B. 2006. Lexicalization patterns in motion events revisited. Ms. ([http://ling.uni-konstanz.de/pages/home/a20\\_11/waelchli/waelchli-lexpatt.pdf](http://ling.uni-konstanz.de/pages/home/a20_11/waelchli/waelchli-lexpatt.pdf))
- Wälchli, B. 2006/2007. Constructing semantic maps from parallel text data. Ms. ([http://ling.uni-konstanz.de/pages/home/a20\\_11/waelchli/waelchli-semmaps.pdf](http://ling.uni-konstanz.de/pages/home/a20_11/waelchli/waelchli-semmaps.pdf))
- Wälchli, B. & Zúñiga, F. 2006. The feature of systematic source-goal distinction and a typology of motion events in the clause. *Sprachtypologie und Universalienforschung (STUF)* 59(3): 284–303. (special issue on *The Lexicon: Typological and Contrastive Perspectives*, T. Leuschner & G. Giannouloupoulou (Eds)).
- Wegener, C. 2006. Savosavo body part terminology. In Majid et al. (Eds), 344–359.
- Wierzbicka, A. 1988. *The Semantics of Grammar*. Amsterdam: John Benjamins.
- Wierzbicka, A. 1990. The meaning of color terms: Semantics, culture, and cognition. *Cognitive linguistics* 1: 99–150.
- Wierzbicka, A. 2007. Bodies and their parts: An NSM approach to semantic typology. *Language Sciences* 29: 14–65.
- Wilkins, D.P. & Hill, D. 1995. When GO means COME: Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6(2–3): 209–259.

## PART II

## Theoretical and methodological issues