Lexical typology: introduction

Maria Koptjevskaja-Tamm &
Stockholm University
tamm@ling.su.se

Ekaterina Rakhilina
HSE
rakhilina@gmail.com

Stockholm
University

About us

Maria Koptjevskaja Tamm (Мария Копчевская Тамм)
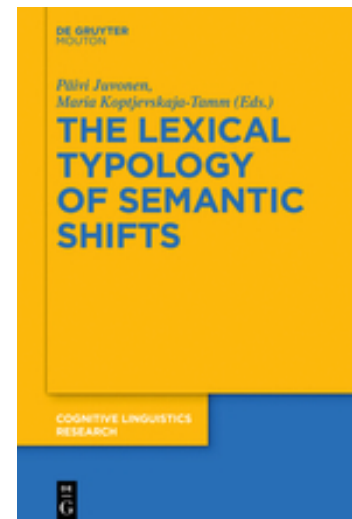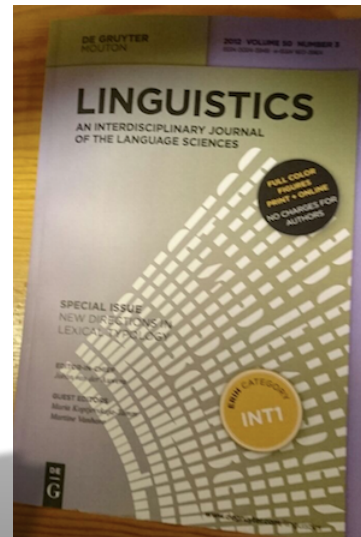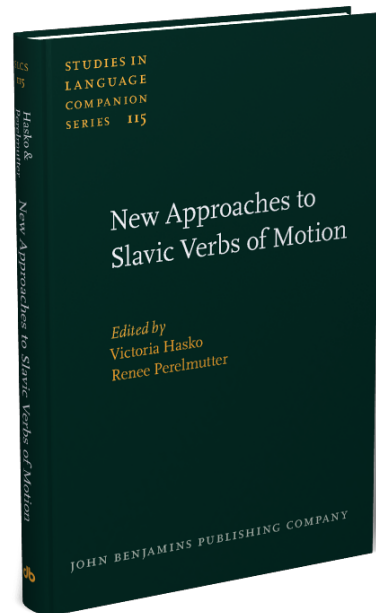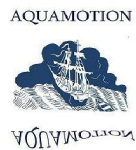[maˈria kɔpˈtʃɛfskaja ˈtam],
Ekaterina Rakhilina (Екатерина Рахилина) [jekateˈrina raˈχilina]

Language diversity (Levinson & Evans 2009)

"The crucial fact for understanding the place of language in human cognition is its diversity. For example, languages may have less than a dozen distinctive sounds, or they may have 12 dozen, and sign languages do not use sounds at all. Languages may or may not have derivational morphology (to make words from other words, e.g., *run > runner*), or inflectional morphology for an obligatory set of syntactically consequential choices (e.g., plural *the girls are* vs. singular *the girl is*). They may or may not have constituent structure (building blocks of words that form phrases), may or may not have fixed orders of elements, and their semantic systems may carve the world at quite different joints… *We are the only known species whose communication system varies fundamentally in both form and content*."

Language diversity (Levinson & Evans 2009)

"The crucial fact for understanding the place of language in human cognition is its diversity. For example, languages may have less than a dozen distinctive sounds, or they may have 12 dozen, and sign languages do not use sounds at all. Languages may or may not have derivational morphology (to make words from other words, e.g., *run > runner*), or inflectional morphology for an obligatory set of syntactically consequential choices (e.g., plural *the girls are* vs. singular *the girl is*). They may or may not have constituent structure (building blocks of words that form phrases), may or may not have fixed orders of elements, **and their semantic systems may carve the world at quite different joints**… *We are the only known species whose communication system varies fundamentally in both form and content.*"

Semantic diversity, ex. 1: body-part terms

English

hand

arm

foot

leg

finger

toe

Russian

ruka

noga

palec

Semantic diversity, ex. 1: body-part terms

English	Italian	Ruman.	Eston.	Japan.	Russ
hand	mano	mina	käsi	ude	ruka
arm	braccio	brat	käsi(vars)	te	
foot	piede	picior	jalg	ashi	noga
leg	gamba				
finger	dito	deget	sõrm	yubi	palec
toe			varvas		

Language diversity and linguistic typology

Linguistic typology – "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)

Typological research takes linguistic diversity as its point of departure. It assumes further that the variation across languages is restricted and aims at discovering the systematicity behind it.

Stockholm University

Kinds of typology: lexical typology

Grammatical typology, syntactic typology, morphological typology, phonetic typology, phonological typology…

Lexical typology – "systematic study of cross-linguistic variation in words and vocabularies, i.e., the cross-linguistic and typological branch of lexicology" (Koptjevskaja-Tamm 2012: 373).

Cf. the "characteristic ways in which language […] packages semantic material into words" (Lehrer 1992: 249)

Our focus – semantically oriented lexical typology.

Stockholm University

Semantic typology

Semantic typology – "the systematic cross-linguistic study of how languages express meaning by way of signs" (Evans 2010:504).

It is orthogonal to the more traditional compartments of typology, such as phonetic / phonological, grammatical or lexical, since meanings are normally expressed by an intricate interplay among signs of various kinds – words, morphological markers, syntactic constructions, prosody, gestures, etc.

Lexical typology: research angle 1

MEANINGS => EXPRESSIONS (onomasiology): how do languages categorize, or carve up particular domains (human body, kinship relations, colour, motion, perception, etc.) by means of lexical expressions?

o E.g., what body-part concepts are encoded as words across languages, what distinctions are made in the systems of body-part terms and what factors underlie them?

o Are languages completely free to "carve up" the domains at an infinite and arbitrary number of places or are there limits on this?

o Are there any universal categories within this domain?

Body

The most universal, basic and crucial human domain.
Extensive research:

      Andersen 1978, Brown 2001, 2005 a/b,

      Wilkins 1996, Majid et al. (eds.) 2006,

      Wierzbicka 2007, etc.

Stockholm University
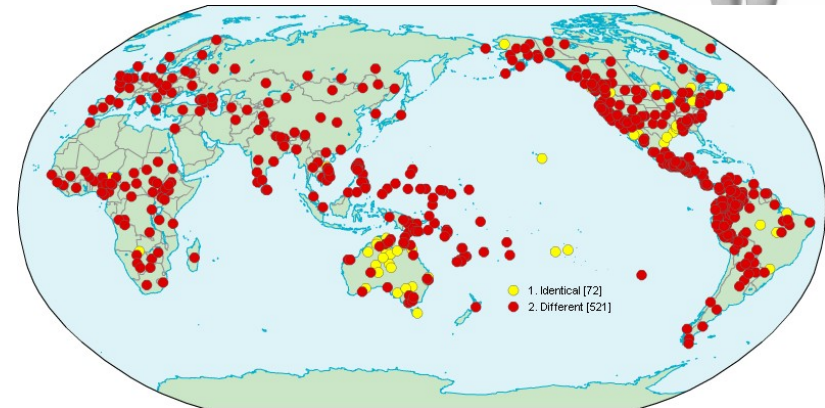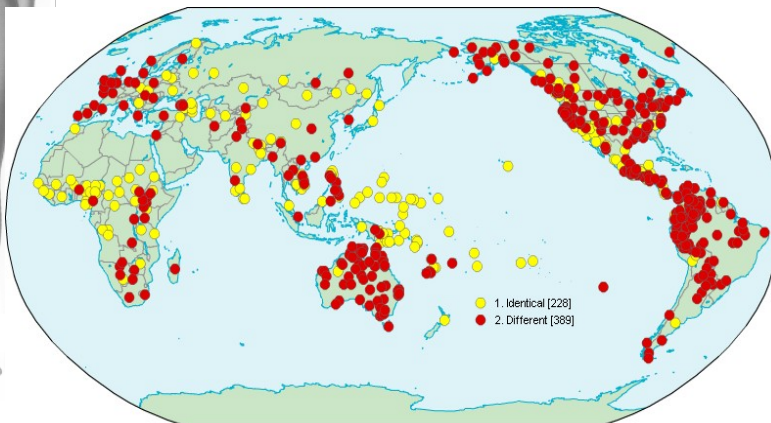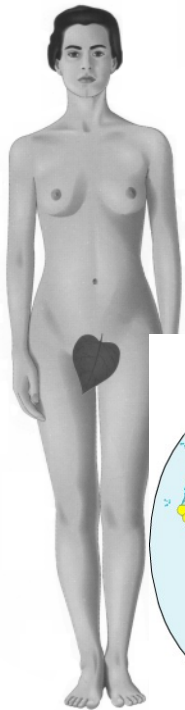
Earlier generalisations

Andersen 1978, Brown 2001

o If both hand and foot are labelled, they are labelled differently.

o If there is a distinct term for foot, then there will be a distinct term for hand.

o There will be distinct terms for BODY, HEAD, ARM, EYES, NOSE and MOUTH

Stockholm University

Universals in the body domain?

Recent research on the BODY (Wilkins 1996, Majid et al. 2006): many of the earlier generalizations have proved to be wrong.

Lavukaleve (Russell islands within the central Solomon islands, Papua) – *tau* 'arm / leg', *fe* 'foot' (Terrill 2006)

'Body': is often the same word as 'person', 'skin', 'trunk'.

Lexical typology: research angle 2

EXPRESSIONS => MEANINGS (semasiology): what different meanings can be expressed by one and the same lexeme or by lexemes synchronically and historically derived from each other (polysemy, semantic shifts etc.)?

o   E.g., what are the possible extensions of body-part terms to other domains?

o   Where from do the body-part terms come?

o   How can their meanings change?

The main focus here is on cross-linguistically recurrent patterns in the relations among the words and lexical items in the lexicon, e.g. semantic motivation (polysemy, semantic shifts) and morphological motivation (word formation patterns).

Stockholm
University

Expressions => meanings: semantic shifts with body parts as source

- o 'Mapping' onto parts of other things
- o Spatial relations
- o Emotions
- o Numbers
- o Possession
- o Reflexive-reciprocal-middles
- o Etc.

Stockholm
University

Bodyparts => spatial relations

(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)
   -'a'ai "in back of" < *'a'at* "anus")

Bodyparts => emotions

(3) Lao (Tai-Kadai) (Enfield 2002: 87)

a.	*aj3-hòòn4*	b.	*caj3-dii3*	c.	*caj3-kaa4*
	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)

gigina-m   [*dal'p*   *a*]

heavy-3PL   stomach.NSG   3M.POSS

'he is worried/sad' (lit. 'his stomach is heavy')

(5) Yélî Dnye (isolate, "Papuan", Rossel island, Papua New Guinea)
(Levinson 2006: 237)

a   *nuu*   *u*   *tpile*

my   throat   his/its/her   thing

'A thing I really like it' (lit. 'My throat its thing')

Stockholm
University

Semantic shifts with body parts as source: universal vs. specific

o Emotions: e.g., 'heart', 'liver', 'stomach', 'throat'

o Spatial relations: e.g. anthropomorphic vs. zoomorphic patterns, 'head' vs. 'trunk' (49% each in Oceania, 38% vs. 60% in Africa)

o Australian Aboriginal languages: animals and plants named after their most salient part, e.g. 'tooth' => 'dog' and 'wild asparagus'

Stockholm
University

Lexical typology: Research angle 3

Lehmann (1990: 163): lexical typology – research which focuses 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).

o E.g., how are body-part concepts lexicalized across languages in terms of word classes?

o Are there morphological peculiarities characteristic for body-part terms?

o What syntactic constructions are used for talking about body parts?

Critical issues in typological research

1. Cross-linguistic identification/comparability of the studied phenomena

2. Methods of data collection

3. Language sample

4. Representation of results

Stockholm University

Crosslinguistic comparability of the studied phenomena and methods of data collection

comparing

like with like

finding a reasonable

level of abstraction

Crosslinguistic comparability of the studied phenomena and methods of data collection

What counts as 'like and like" is often dependent on the research object and goal. But crucially cross-linguistic identification of phenomena should involve theory-neutral or framework-neutral definitions and concern "observable phenomena that pattern interestingly in the world's languages" (Nichols 2007: 231).

Stockholm University

Back to the generalizations on body parts: Why these differences?

Earlier research – predominantly dictionary based

Recent research on the BODY (Majid et al. (eds.) 2006, Majid et al. (eds.) 2015): systematic elicitation of primary data based on the same guidelines

Methodology is important!

Stockholm University

What is a hand? Variation in dictionary definitions

hand

- the part of the human body attached to the end of the forearm, including the wrist, palm, fingers, and thumb [YourDictionary.com]

- The terminal part of the arm beyond the wrist, consisting of the palm and five digits, forming the organ of prehension characteristic of man. [OED online]

Stockholm University

What is a hand? Variation in dictionary definitions

hand

- the part of the human body attached to the end of the forearm, including the wrist, palm, fingers, and thumb [YourDictionary.com]

- The terminal part of the arm beyond the wrist, consisting of the palm and five digits, forming the organ of prehension characteristic of man. [OED online]
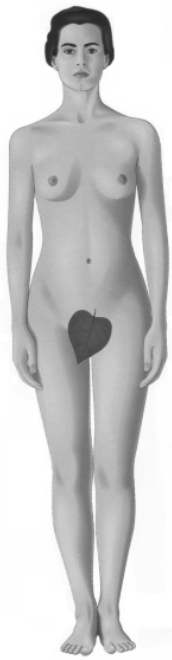
Stockholm University

The body elicitation guidelines

Enfield, Nicholas, 2006, Elicitation guide on parts of the body.
In Majid, Asifa, Nicholas J. Enfield and Miriam van Staden
(eds.) Parts of the body: Cross-linguistic categorisation.
Language Sciences, 28(2-3), pp. 148–157

http://fieldmanuals.mpi.nl/

Stockholm
University

Denotation-based lexical typology

Brent Berlin and Paul Kay in the 1960s: Munsell colour charts – a number of extra-linguistic contexts for capturing possible distinctions within the colour domain (organized according to the three dimensions of hue, value/lightness, and chroma/color purity), which enables comparison of denotational ranges of colour terms both within one and the same language and across languages.

A major part of research on lexical typology has been conducted on domains whose denotation / extension lends itself easily to description / stratification by means of simple behaviouristic procedures.

Stockholm
University

The "Nijmegen method" of semantic typology

o   The "Nijmegen method" of semantic typology uses standardized stimuli – sets of pictures, videoclips and films – for collecting data on a number of cognitive domains directly in the field (cf. http://fieldmanuals.mpi.nl/).

o   Each set covers a shared denotational grid allowing systematic comparisons of semantic distinctions potentially relevant for the domain and may be used under different elicitation conditions, including games.

Stockholm
University

Etic definitions

the meanings of linguistic expressions = sets of uses, 'etic definitions'.

Stockholm
University

Example – the "CUT and BREAK" project (Majid and Bowerman 2007

The domain: state changes involving some kind of separation in an object, "separation in the material integrity of objects", which is in the normal case, but not always, explicitly caused by an agent.

To cut hair (with scissors), to cut bread (with a knife), to slice a carrot, to break a stick over the knee, to break a vase, to tear a cloth into two pieces, etc.

Stockholm University

The main issues

o   To what extent languages will agree or disagree in their categorization of this domain, i.e. in which experiences will be encoded by one and the same expression or by different expressions?

o   To what extent is linguistic categorization universal or language- and culture specific?

Stockholm University

Methodology: data collection and the sample

61 videoclips, 28 languages

(1) Tear cloth into two by hands

(27) cut hair with scissors

(39) smash a pot with a single blow of hammer
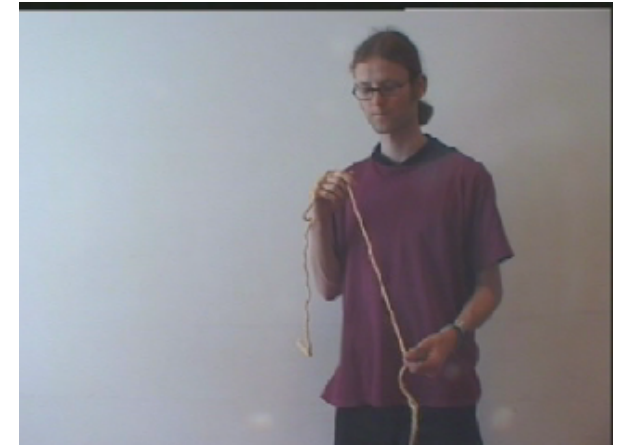
Stockholm University

Methodology (continued)

(13) cut rope stretched between two tables with a blow of an axe

(35) break yarn into many pieces with fury

(49) cut rope in two with knife

Stockholm
University

The results

o An impressive cross-linguistic variation in how languages encode and categorize the domain.

o However, this cross-linguistic variation operates within a constrained space, that may be defined in terms of cross-linguistically important dimensions, contrasts, or parameters.

We start with a few examples of cross-linguistic variation and end up with generalizations on how it is constrained.

Stockholm
University

An exotic example: Yélî Dnye (Papuan, Rossell island, Levinson 2007)

Three core verbs. The crucial notion – grain, or aligned fibers, cf. wood, cloth, leaves vs. stone, pottery.

châpwo
SEVER AGAINST THE GRAIN:
1. cut with instrument
2. chop with hand

chaa
SEVER WITH THE GRAIN:
1. split with instrument
2. tear by hand without instrument

pwââ
BREAK (divide incoherently, regardless of grain)
1. with sharp instrument
2. with blunt instrument
3. with hands

Stockholm University

Yélî Dnye (Papuan, Rossell island, Levinson 2007): pwââ

No explicit instrument

a. *pyââ*   *ngê*   *d:ââ*   *dê*   *pwââ.*
   woman   ERG   clay.pot   3SIMMPASTPUNCT   broke.trans
   'The woman broke the pot.'

An explicit instrument

   pi   *ngê*   *hammer*   *ngê*   *k:aa*   *dê*
   person ERG   hammer   INSTR   taro   3SIMMPASTPUNCT
   pwââ.
   broke.trans
   'The man broke the taro with a hammer.'

Stockholm
University

With or against the grain

chaa
SEVER WITH THE GRAIN:
1. split with instrument
2. tear by hand without instrument

châpwo
SEVER AGAINST THE GRAIN:
1. cut with instrument
2. chop with hand

Instrument is irrelevant, probably reflecting the culture of a century ago, with no metal instruments, only blunt basalt axes.

Stockholm University

Mandarin (Chen 2007)

The typical structure of cutting and breaking expressions – a resultative verb compound (RVC): all in all 43

Ta1	**qie1-duan4**		le	shen2zi
he	cut.with.single.blade-be.broken		PFV	rope

'He cut the rope.'

action verb   result verb

Stockholm
University

Mandarin action verbs (Chen 2007)

o CUT with one blade

o CUT with scissors

o BREAK with a hammer-like instrument

o TEAR

o BEND / PULL ON A LINEAR OBJECT WITH HANDS

Both instrument and manner are important

Stockholm
University

Chen 2007

Table 1. *Mandarin 'cutting with single blade' verbs in the elicited data*

| Verbs	Glosses
qie1	'do cutting with a single blade or blade-like instrument'
kan3	'do cutting with a single blade or blade-like instrument with force'
duo4	'do chopping, dicing, repeatedly'
zhan3	'do chopping, cutting cleanly' (often in literary use)
pi1	'do hacking, cutting with force and usually into halves, cleave'
po4	'dissect, cut carefully'
ge1	'do cutting with a single blade or single-blade-like instrument slowly, duratively, back and forth'
zao2	'do cutting with a chisel'
chuo1	'do cutting with a sharp pointed instrument'
ju1	'do cutting with a saw in a sawing manner, back and forth'

Stockholm
University

Is the cross-linguistic variation not constrained at all (Majid et al. 2007a)?

Multiple dimensional scaling => Dimensions underlying the cross-linguistic variation

Stockholm
University

o   Dimension 1: the relative predictability of the locus of separation in the object acted upon.

o   Cf. slicing carrots with a small knife, smashing a plate with a hammer, and chopping off a branch with an axe.

While all languages distinguish between the events at the extremes, the events in-between may be treated differently by different languages.

Stockholm
University

o Dimension 2 (within low predictability): smashing a rigid object like a plate, pot, or carrot with a sharp blow vs. snapping a long object like a stick or a carrot into two pieces between the hands or over the knee.

o Dimension 3: tearing events vs. all the others

Stockholm
University

Stockholm
University
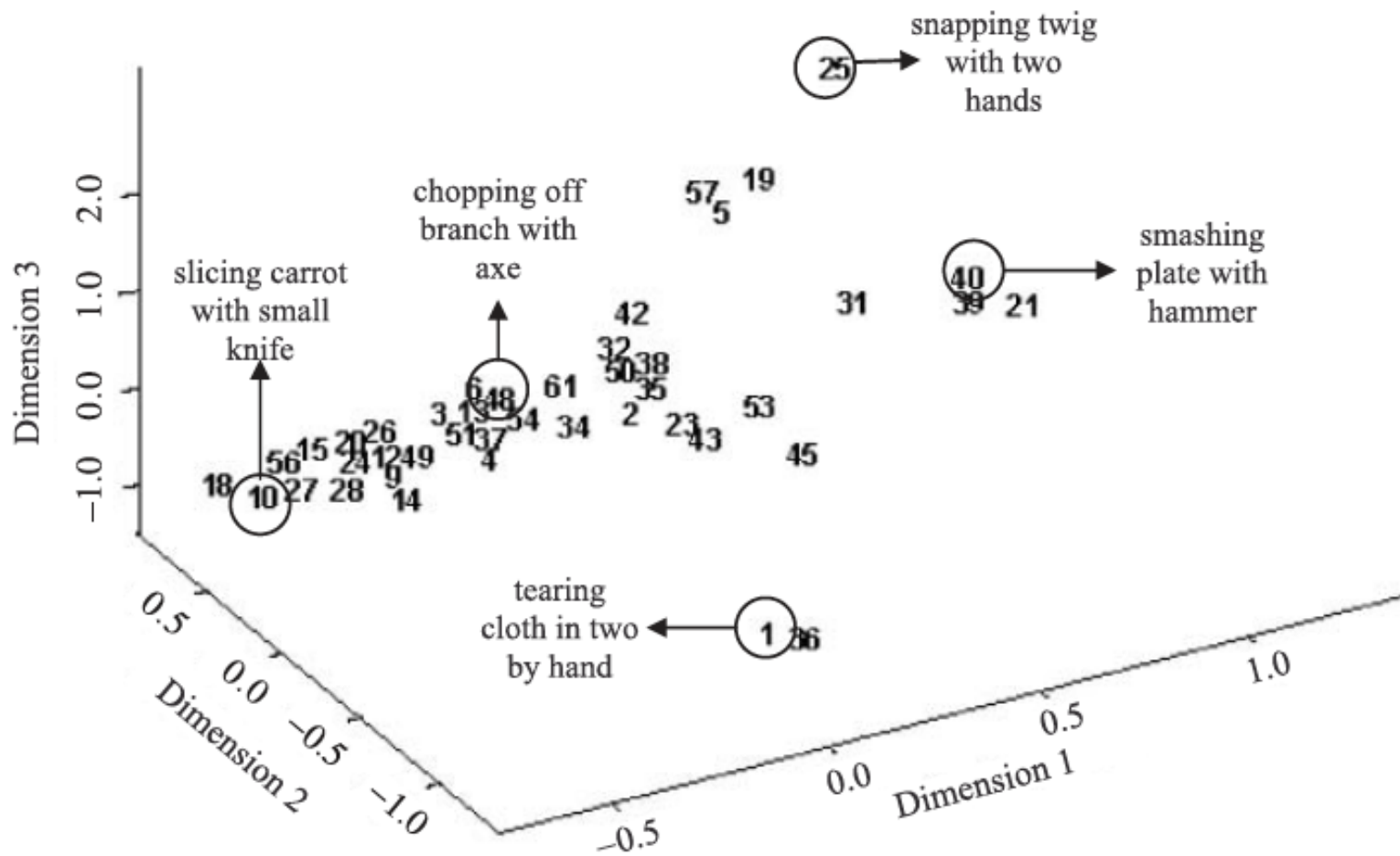
Denotation-based methods: general

o Different methods for different domains.

o Quine's "Gavagai" problem: how does a learner know what an observed instance of a word used in context refer to?

o "Denotation-based" definitions work differently well for different kinds of situations. Taste, temperature, sound etc. – we have to work out further procedures. Emotions, thoughts, pain etc.?

Stockholm University

Denotation-based methods: general

o Decontextualization: different for different techniques (cf. retelling a film for someone who has not seen it or exchanging verbal instructions during a game vs. describing a series of disconnected videoclips or naming colour chips).

Stockholm
University

Denotation-based methods: general

o Not always sufficient attention to the linguistic behaviour of the linguistic expressions. Lucy (1997): the mainstream tradition of research into colour terms across languages does not presuppose any deeper linguistic analysis of these terms. "Articles surveying terms in a dozen or more languages never mention anything about those language, or even about the structural value of the terms. You do not need to know anything about languages or linguistics at all to read this literature or even to conduct research within the tradition" (ibid.: 330).

Stockholm
University

MOSCOW LEXICAL TYPOLOGY GROUP

"The frame method" of lexical typology

Primarily developed within the Moscow Lexico-Typological group (MLexT)

1. Wide range of linguistic resources

2. Theoretical background: linguistic behavior of lexical items (in the traditions of Moscow semantic school)

3. Frequent situations (frames) as "primitives" for cross-linguistic comparison of lexical domains

4. Frame as sources for semantic extensions (typology of metaphors)

Stockholm
University

1. Linguistic resources for lexical typology

- **No ideal dictionaries:**

- Lexicographic data taken from dictionaries as standard open resources for lexical semantics is never complete, consistent or compatible → it is never sufficient for the detailed cross-linguistic comparison.

- Lexical typologists use a wide range of resources:

➢ all kinds of dictionaries

➢ corpora

➢ evidence from native speakers (they fill in the questionnaires, check the examples, etc.)

Stockholm University

2. Theoretical background: linguistic behavior of lexical items

- The theory is supported by the tradition of Moscow semantic school (MSSchool) and could be called "combinatorial lexical typology" (Koptjevskaja-Tamm et al. 2015).

- Semantic differences in the domain are obtained through comparison of close synonyms (Apresjan 2000)

- Translation equivalents in languages of the sample can be treated / viewed as close synonyms → we compare their contextual behaviour.

Stockholm University

3. Frames in cross-linguistic comparison

- Frames (= typical situations characteristic for the domain) serve as entries for typological questionnaires.

Oscillation

pendulum

tree

old fence

curtain in the wind

Typology: two goals

- Universal list of frames for each lexical domain
- Patterns of their colexification

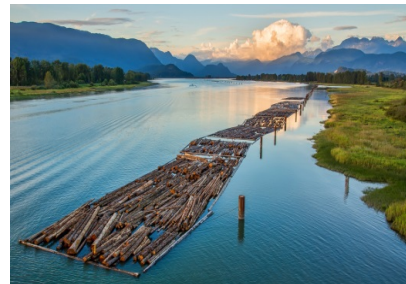$\Rightarrow$ Cognitive strategies of combining different frames by the same lexeme

Aqua-motion. Four main frames:

'swim'   'sail'   'drift'   'float'

Stockholm
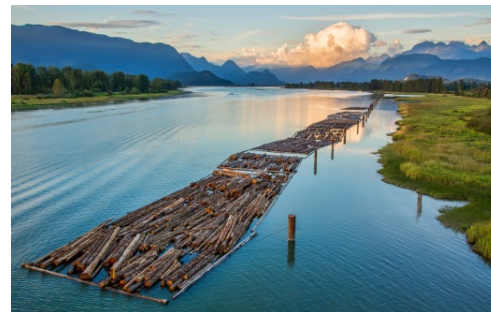University

Colexification: dominant system (a single word for all frames)

- Bulgarian: *pluvam*

swim   sail   drift

float
Stockholm
University
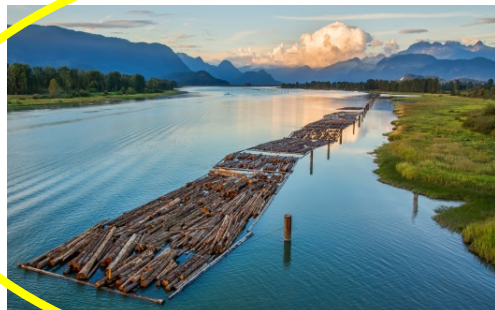
Colexification: classifying systems

Tamil

swim   ← – –   sail   – – →   drift

Colexification: classifying systems

- HINDI

swim

float

sail

drift

Patterns of colexification

TAMIL

HINDI

4. Typology of metaphors

- Each frame has its own way of semantic development and is a source for its own extensions

To squeeze one's way through the crowd
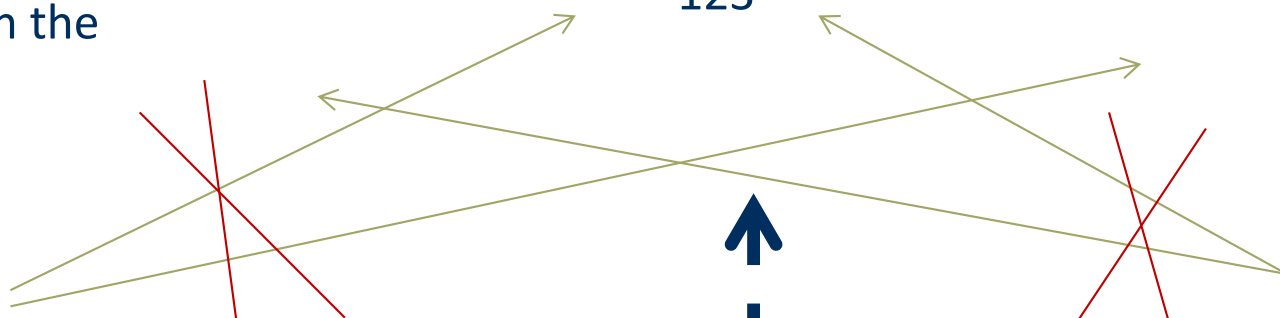
123

Hesitation
Non stability

swim

drift

float

Moscow Lexical Typology Group typological projects:

- Verbs of aqua-motion
- Pain metaphors
- Sound metaphors
- Verbs of rotation
- Verbs of oscillation
- Cutting & breaking
- Sitting & standing
- Physical qualities ('sharp', 'wet', 'soft', 'even'…)
- Falling

Stockholm University

In preparation:

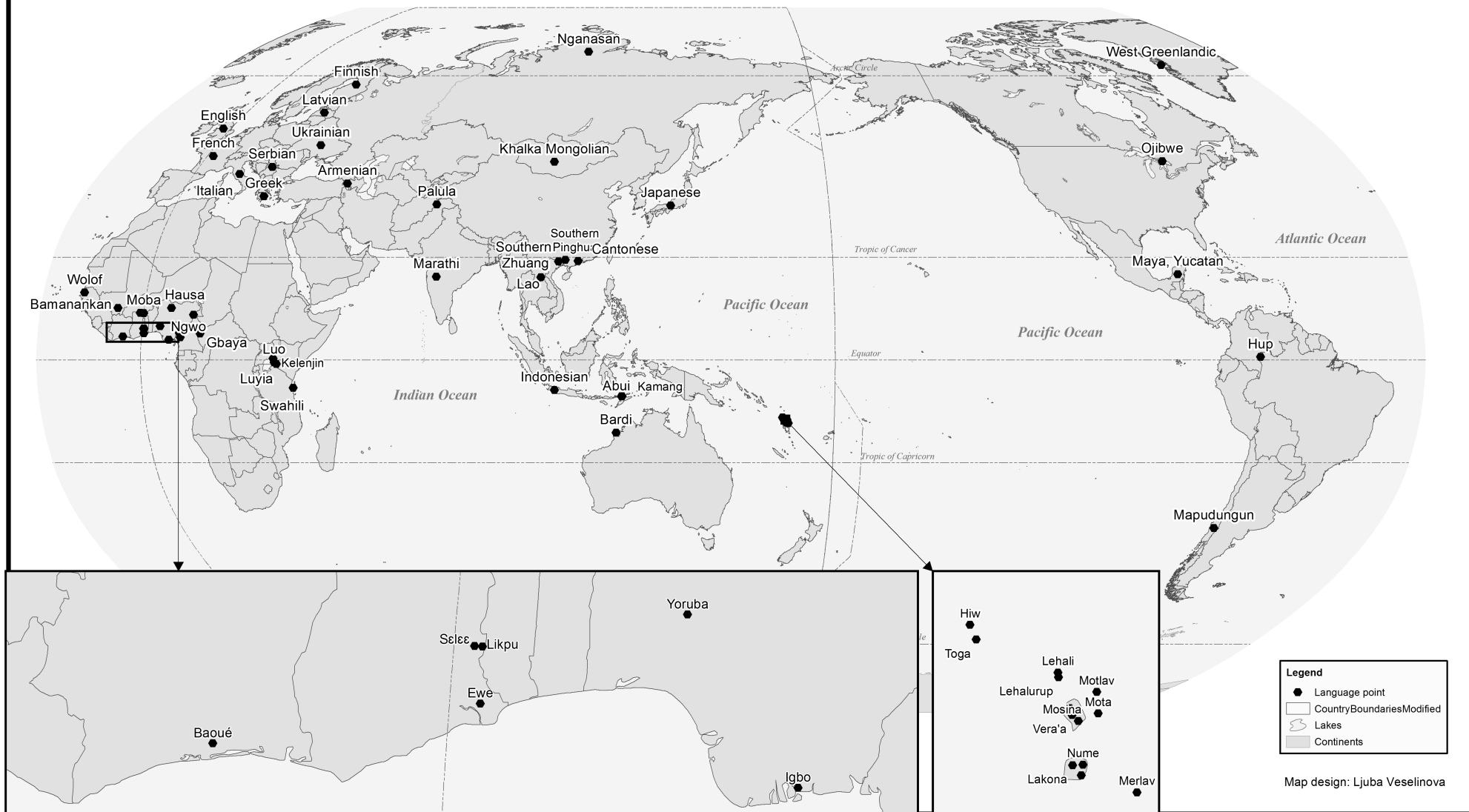- "Linguistics of qualities" (2017)
- "Verbs of falling" (2017)

The linguistics of temperature

❑ Koptjevskaja-Tamm (ed., 2015), "The linguistics of temperature" (Benjamins):

✓ 27 chapters with detailed descriptions of 50 languages

✓ 3 cross-linguistic chapters (not all of the languages are used here)

University

Overview of the languages investigated in this volume

Languages shown on the map include:

Nganasan, Finnish, Latvian, English, French, Ukrainian, Serbian, Armenian, Italian, Greek, Khalka Mongolian, Palula, Japanese, Marathi, Southern Zhuang, Southern Pinghu, Cantonese, Lao, Maya, Yucatan, Wolof, Bamanankan, Moba, Hausa, Ngwo, Gbaya, Luo, Kelenjin, Luyia, Swahili, Indonesian, Abui, Kamang, Bardi, Hup, Mapudungun, West Greenlandic, Ojibwe

Inset (West Africa): Yoruba, Sɛlɛɛ, Likpu, Ewe, Baoué, Igbo

Inset (Pacific): Hiw, Toga, Lehali, Motlav, Lehalurup, Mota, Mosina, Vera'a, Nume, Lakona, Merlav

Legend
- • Language point
- CountryBoundariesModified
- Lakes
- Continents

Map design: Ljuba Veselinova

Frames of temperature evaluation

TACTILE temperature, or "touch-temperature"
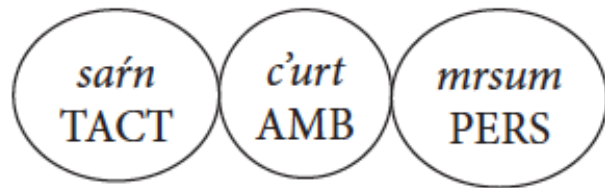
(1a)   *The stones are hot*

AMBIENT temperature

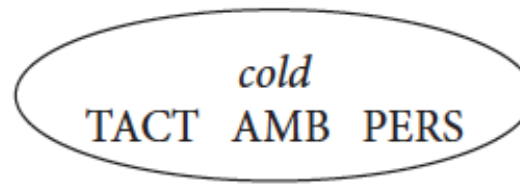(1b)   *It is hot here*

PERSONAL-FEELING temperature

(1c)   *I am hot* (because the room is too heated, because I have been running…)

Cf. Plank 2003, FrameNet *(http://framenet.icsi.berkeley.edu/,* Goddard and Wierzbicka 2007

Figure 1. 'Cold' in Armenian, English, Palula (Indo-European), Japanese (Japonic) and Kamang (Timor-Alor-Pantar) (TACT = TACTILE, AMB = AMBIENT, PERS = PERSONAL-FEELING, QR = QUASI-REFERENTIAL, NR = NON-REFERENTIAL)

Language sample in lexical typology

o The language sample for lexical typology is quite limited as compared to grammatical one

➤ Grammar: 200-400 languages

➤ Lexicon: 15-50 languages

We find that even a study of 15 languages allows for some non-trivial generalizations

- The sample for lexical typology can contain related languages.

- Lexical changes are more rapid than those in grammar or phonetics. Some of them could be registered within one and the same generation of speakers. Related languages (and even distant dialects) display typologically relevant lexical violations.

- Polish, Ukranian & Russian, Spanish, French & Italian, German, Dutch & English can be members of the same sample

Stockholm
University

Some of the best researched lexical fields/ semantic domains

✦ Body

✦ Kinship

✦ Colour

✦ Perception

✦ Motion (both inspired by Talmy, but also deictic verbs – Ricca, multiple questions – Wälchli, aqua motion – Maisak & Rakhilina (Moscow)

✦ Dimension

✦ Posture (Ameka & Levinson 2007)

✦ Cut / Break (Majid et al. 2007)

✦ Smell (Asifa Majid, MPI, Nijmegen)

✦ Pain

Lexical typology

Stockholm University

Thank you!

Stockholm
University