CLIPP

Christiani Lehmanni inedita, publicanda, publicata

titulus
Converse categorization strategies
huius textus situs retis mundialis
http://www.christianlehmann.eu/publ/categorization.pdf
dies manuscripti postremum modificati
17.04.2012
occasio orationis habitae
Workshop on Lexical Typology of the 7th Meeting of the ALT, Paris, 23-27 Sept 2007
volumen publicationem continens
Koptjevskaja-Tamm, Maria & Vanhove, Martine (eds.), *New directions in lexical typology*. Linguistics 50/3 (Special Issue)
annus publicationis
2012
paginae
467-494

Converse categorization strategies

Christian Lehmann

University of Erfurt

Dedicated to Hansjakob Seiler on his 90th birthday.

Abstract

The categorization alluded to in the title is the assignment of a class of concepts to a lexeme class and/or a syntactic category. The purpose of the paper is to establish converseness of strategies of categorization among languages as a typological concept. It is argued that, quite in general, coding strategies in a given functional domain may be oriented in opposite directions across languages. Particular attention will be paid to the relationship between basic/lexical categorization and derived/syntactic categorization.

A particular kind of converseness is produced by the alternative of basically lexicalizing some concept in grammatical category C_i and transferring it into category C_j by derivational or grammatical operations, or vice versa. The chief empirical domain to illustrate the principle is the categorization of dynamic relational concepts as verbs vs. non-verbs, the latter paired with prominence of light-verb constructions in the grammar. A couple of other functional domains susceptible of the same kind of analysis are analyzed more summarily.

Whenever the elements of a certain conceptual field or functional domain are uniformly lexicalized in some particular category, this is typically coupled with a regular operation of recategorization into its complementary category. In such cases, both the basic category assignment and the presence of the operation shape the structure of sentences and of texts in the language.

1 Introduction[1]

The *primum datum* of linguistic typology is variation, the elementary observation is: some languages do it this way, other languages do it another way. It is the linguist's task to go beyond this trivial observation by systematizing the variation and discovering the underlying principles. The basic alternative for a language is often between coding some particular distinction or leaving it to inference. The temporal relation of the designated situation to speech-act time as coded by tense is a relevant example. However, in certain functional domains, constraints are tighter. Certain kinds of categorial information are so basic that they are generally coded at some level. This concerns, above all, the parts of speech. Now while the task of at all assigning some conceptual class to some structural category can hardly be dodged, such distinctions as between noun, verb, adjective etc. are not made by all languages alike, so that languages differ both in the particular categories and in the level at which categorial information is assigned (cf. Lehmann 2008). For instance, something may be in the category of nominal expressions either by belonging to the word-class 'noun' or by having been nominalized, by a process of lexical derivation or by some syntactic operation.[2]

Quite in general, categorization strategies in a given functional domain may instantiate opposite possibilities among languages. This applies, above all, to the relationship between basic/lexical categorization and derived/syntactic categorization. A particular kind of converseness is produced by the alternative of lexicalizing some concept in grammatical category C_i and transferring it into category C_j by derivational or grammatical operations, or vice versa. This paper will explore such mirror-image relationships. In doing so, it will be a contribution to the basic question "What cross-linguistic patterns are there in lexicon-grammar interaction?" (Koptjevskaja-Tamm 2008:6).

This kind of converseness between the categorization and recategorization strategies of two languages will be illustrated with an in-depth case study of light-verb constructions in Persian and German. Subsequently, the general applicability of the conception will be plausibilized by a more summary review of a couple of other empirical areas. The data are drawn from the published literature. The potentially novel contribution here is their typological comparison and the empirical substantiation of typological converseness relations, to be defended at a general methodological level in section 5.

A terminological clarification is necessary concerning the term *lexicalization*. It is here used in two related senses: 1) For a concept to be lexicalized in a specific language means for it to be associated with a lexeme, which implies – importantly in the context of this paper – assigning it to some lexical category. In that sense, an entire set of concepts may be

---

[1] Paper read at the Workshop on Lexical Typology of the 7th Meeting of the ALT, Paris, 23-27 Sept 2007, at Dulzon's Readings 25, Tomsk, 25-29 June, 2008 and the Research Centre for Linguistic Typology, La Trobe University, 10 November, 2010. I thank two benevolent-severe anonymous reviewers for their helpful criticism and suggestions, which have shaped the paper quite a bit. As a consequence, the two bear full responsibility for any remaining errors.

[2] This idea was already expressed in Benveniste 1957: 222: les unités complexes de la phrase peuvent, en vertu de leur fonction, se distribuer dans les mêmes classes de formes où sont rangées les unités simples, ou mots, en vertu de leur caractères morphologiques. (the complex units of a clause may, by virtue of their function, be distributed in the same form classes in which simple units, or words, are arranged by virtue of their morphological features.)

lexicalized in a particular category. This is the sense in which the term is used, *i.a.*, in Talmy 1985. In the present context, basic lexicalization plays a prominent role; this is lexicalization in the form of a (categorized) root, providing, thus, at the same time a basic categorization of a concept. 2) For a linguistic sign to be lexicalized (to a high degree) means for it to be inaccessible to compositional rules and, thus, to be idiomatic. In this latter sense, lexicalization has implications for the mental lexicon which, however, are of no concern to us. In the sequel, each occurrence of the term will be disambiguated by the context.

2 Complementarity and converseness across languages

In universals research, a major breakthrough was achieved when Joseph Greenberg (1963) used the Jakobsonian notion of **implicational relationship** between units of a linguistic system and applied it at the typological level. It then became clear for the first time that a language universal need not be a property of the language system that appears in all languages. Instead, there was a logical relationship between two different properties of the language system concerning their distribution across languages. This was a more abstract kind of language universal that pointed to some principle regulating the buildup of linguistic systems.

An implicational generalization formulates a certain distribution of two properties p and q across languages in terms of propositional calculus. The distribution in question is shown in the four rows of T1, which is the familiar truth-value table for the conditional. Among the many other ways that two properties may be distributed across languages, complementary distribution has been of special interest. Two properties p and q are in complementary distribution over some domain iff p does not occur in the contexts that q occurs in, and vice versa. In terms of propositional logic, that is the relation of **contravalence**, For a comparison of implication with complementary distribution in terms of propositional calculus, the truth value tables are here juxtaposed:

T1 Conditional

p	**q**	**p → q**
t	t	t
t	f	f
f	t	t
f	f	t

T2 Contravalence

p	**q**	**p × q**
t	t	f
t	f	t
f	t	t
f	f	f

The contravalence pattern might, in principle, apply to the cross-linguistic domain just as implication has been applied to it ever since Jakobson and Greenberg. The first one to propose this was Hansjakob Seiler in his contribution to the International Congress of Linguists 1972. His example then was the functional domain of possession. Seiler 1998 and 2010 present more recent attempts to come to grips with complementary relationships.

I will illustrate the point with an example that, although falling short of the strictest requirements, comes close to the idea formalized in T2. Consider the statement: Every language has either prepositions or postpositions. With respect to the four rows of T2, this has the following implications:

- Some languages have prepositions, but no postpositions. French and Arabic are relevant examples.
- Some languages have postpositions, but no prepositions. Relevant examples include Turkish and Japanese.
- No language possesses both prepositions and postpositions. This does not hold as it stands. The WALS (Haspelmath et al. 2005) lists 52 languages without a dominant order for adpositions in a sample of 1074 languages,[3] with O'odham, Pashto and Somali among them.
- Every language has adpositions. This may again be false, depending on how wide our definition of adposition is (cf. DeLancey 2005). The WALS lists 28 languages without adpositions, among them Blackfoot and Dyirbal.

While the individual data await finer analysis, the generalization stating a complementary distribution of prepositions and postpositions would be true as a statistical tendency, with 7% of exceptions in the WALS sample.[4] If we wanted to avoid the exceptions, we might refine the generalization by narrowing down the implicans, for instance like this: If a language has adpositions, then either prepositions or postpositions are the dominant strategy. However, the grammar of adpositions is not at stake here; they just serve as a simple example.

Why would complementary distribution be important? Distributions of linguistic properties, if principled rather than fortuitous, are indicative of certain inner relationships between the entities concerned. These are of a different kind for implicational and contravalent relationships:

- In implication, the two properties are not on the same level; one is more important for the language system than the other. Consider the example: If a language has front rounded vowels, it has back rounded vowels. The implicatum is the basis for the implicans. Thus, the basic rounded vowels are the back ones; the front ones are more complex. So much can be inferred from the distribution itself. The next step is now a search for the factors that make back rounded vowels relatively basic, but front rounded vowels relatively complex.
- In contravalence, two properties are alternatives for a language system. They are on the same level and equally important and useful for a language to have. Two units of a language system that are in complementary distribution are isofunctional. The same would be true for linguistic properties distributed complementarily across languages. The domain in question is structured in such a way that there is just a binary alternative such that every language opts for one of the two possibilities.

In the example of the adpositions, what is at stake is their order relative to a point of reference (their complement NP). With appropriate provisos, that is a binary parameter with the values 'before' and 'after'. In such cases, either of the two possibilities is fully sufficient for a language system; no loss is involved in not having the other alternative available. Nor is there a regular way of generating prepositions from postpositions, or vice versa, i.e. of transposing adpositions.

---

[3] It would both be of general interest and reduce the number of counterexamples to the third implication if the criterion of dominance could be refined. Dryer (2005:7) seems to apply a criterion of basicness and frequency, by which even a very small set of highly grammaticalized prepositions besides a large set of postpositions would exclude the categorization of the language as dominantly postpositional (and vice versa). If the criterion were productivity, instead, then such a language would be clearly postpositional.

[4] I assume the seven inpositions can be subsumed under either pre- or postpositions.

Clean complementary distributions of features across languages are rare if at all existent. T2 presupposes some function that must be fulfilled by languages and can only be fulfilled in either of two ways. Such states of affairs are rare. Even if one of the two solutions is fully sufficient, what could prevent a language from acquiring the alternative solution, too, clashing thus with the first row of the truth-value table?

In what follows, we will focus on cases that come as close as possible to a complementary distribution: In a given functional domain, there is a principal alternative between solutions p and q. At a basic level, a language may opt for alternative p; but since q is also useful, it has an operation to convert p into q, which then is at the next higher level of structural complexity. The opposite goes for a language that opts for q at the basic level. Those languages that opt for one of the two converse solutions therefore typically develop strategies that are mirror-images of each other. On the other hand, since both alternatives are useful to have, there is little to force a language to be consistent in its basic choice. Consequently, such categorizations and associated strategies are generally not in a clean complementary distribution across languages. For the same reason, the case studies presented in the following two sections will not involve any large-scale cross-linguistic comparison and instead abide by contrasting two languages each which display the **converse relation** in question. A mass comparison in any of the functional domains illustrated here would certainly bring out a continuum between the opposite poles.

The next section provides a rather detailed comparison of the role of light-verb constructions in Persian and German, arguing that their structural parallelism is deceptive because in Persian, the light-verb construction is a basic construction, while in German it is at a higher level of structural complexity. Section 4 reviews more briefly two further domains in which languages use converse strategies, viz. basic transitivity vs. intransitivity of verb stems and adpositions vs. relational nouns coding spatial regions, and finally recalls a few more domains that have been touched upon in earlier research and would deserve closer analysis from the point of view suggested here. Section 5 explicates methodological aspects of this kind of typology. The conclusion points to some aspects of theoretical interest of opposite solutions to cognitive and communicative problems in languages.

3 Light verb constructions in Persian and German

This section is devoted to the part-of-speech categorization of dynamic relational concepts (DRCs), i.e. of types of processes and events. These are occasionally called ‘verbal concepts’. However, that term fails both by being eurocentric and by trying to capture the essence of a semantic entity by a structural criterion (cf. fn. 19). The point of this section is precisely that languages differ in categorizing the bulk of DRCs either as verbs or as something else.

3.1 Light verbs and verb completors

Most if not all languages base their coding of DRCs on the category of verb, lexicalizing at least a distinguished subset of them as verb roots. On this common basis, divergent strategies of coding the thousands of DRCs are pursued. One basic alternative is between abiding by a small, closed class of verbs and employing syntactic means to code complex DRCs, vs. working with a large, open class of verbs which may include a large number of verb roots and a set of derivational operations at the stem level to feed that inventory. The latter alternative is

well known from SAE languages and will not be further differentiated here. The former alternative has first been described for languages of Australia and Papua New Guinea, which may have as few as 10 verbs (Schultze-Berndt 2003:147). This basic choice allows for a variety of more specific strategies. One is to code complex DRCs by verb serialization. A well-documented example of this is Kalam (Trans New Guinea) (Pawley & Lane 1998). Another strategy is to entrust a different lexical category with the bulk of DRCs. The core of a dynamic situation is then coded by a binary construction whose head is one of that closed set of verbs and whose dependent is one of that open class of non-verbal elements. In descriptions of Australian languages, this dedicated category has been variously called preverb (Schultze-Berndt 2003), converb, adverb, verb adjunct (Pawley & Lane 1998). Outside Australia, similar categories have been dubbed unmarked stem (Sakel 2007) and even classless word (Karimi-Doostan 2006). Both the individual neologisms and their proliferation testify to the lack of systematic typological research in this area. In the present treatment, a category whose primary function it is to combine with verbs to form complex verbal expressions that code complex DRCs and which does not coincide with nominal or adverbial categories of the same language will be called **verb completor**.

Where there is a closed class of verbs, these may articulate the domain of dynamic situation cores into a small set of very generic types like rest, motion, transfer, production etc. Where that is the case, and to the extent that the combination of a verb and a verb completor is compositional, the verb may be conceived of as classifying DRCs and be called a verbal classifier (McGregor 2002, among others). In a variant of this strategy, the binary construction morphologizes and the verbs become verbalizers (called verbness markers in Sakel's (2007) account of Mosetén). Again, for a given verb completor, there may be a small paradigm of grammaticalized verbs or verbalizers which alternatively combine with it in a regular way, coding diathetic or aktionsart variants of the DRC, as in E5 below. These verbs do not so much classify verb completors, but rather function as operators in regular recategorization operations. Finally, the combination of the verb and the verb completor may not at all be compositional, and instead each such collocation may be lexicalized as a phrasal verb, as in E1 below. Several of these variant strategies may co-exist in one language. In English, for instance, *pay attention (to X)* may have the same meaning as *attend (to X)*. To the extent that one of the strategies prevails in a language, it may characterize its type.

In SAE languages, verbal constructions whose structural head is a rather empty verb taking a dependent which semantically enriches the situation core rather than representing any of the participants are known as light verb constructions. A **Light Verb Construction** (LVC) is a construction of the structure shown in S1 (sequential order of A and B is irrelevant),

S1 General structure of the Light-Verb Construction

$[\ \ldots\ [\ [A]_{C}\ [B]_{V}\ ]_{LVC}\ ]_{VP}$

where B is one of a set of verbs of generic function and meaning, and C is the syntactic category of the component that contributes the bulk of the lexical meaning. B is called a **light verb**. A functions as the **inner dependent** of the LVC. The LVC is the core of a VP, which has valency and admits of adjuncts just like a simple full verb. The three dots represent complements and adjuncts of the LVC.

If more precision in the internal structure of S1 is sought, language-specific differences come into play. In an SAE language such as German, serving as an example in the next subsection,

C is no part of speech of its own and instead any of the syntactic categories that can constitute a dependent of the verb B. In the syntactically regular cases to be analyzed below, that dependent is a complement, thus, either an NP or a PrepP. This is different in Persian, the language used for contrast below, as well as in other languages with a restricted set of verbs. Those languages may instantiate C by a particular part of speech, the verb completor, abbreviated VC, which need not be a complement of the verb. It may instead be a modifier; or the language may not even distinguish between these two kinds of verbal dependents.

The point of the following comparison of German and Persian LVCs is that the two languages make opposite choices concerning the part-of-speech categorization of the bulk of DRCs, which is responsible for the different prominence of LVCs in the two language systems. The functions fulfilled by LVCs, however, are partly similar in the two languages: LVCs are partly compositional, partly idiomatized, and to the extent they are compositional, LVs function as operators of voice and valency change and of aktionsart. In neither language do they have a classificatory function. These functional similarities will be illustrated in passing in order to show that the opposite choices are actually made in a well-delimited functional domain.[5]

3.2 Light-verb constructions in German

A subset of LVCs in German, illustrated in E1, are idiomatized.

E1 a. hops gehen
GERMAN hop(PTL) go:INF
'to die'

b. hops/hopp nehmen
hop(PTL) take:INF
'to arrest'

These will not occupy us any further. Suffice it to note that the inner dependent in E1 is not a (simple or cased) NP, but an ideophone. The complementary subset of LVCs forms rather regular and productive groups, as those in E2 (cf. Lehmann 1991, §3.5).

E2 a. das Programm kommt zur Ausführung
GERMAN the program comes to:the execution
'the program is executed'

b. sie bringt das Programm zur Ausführung
she brings the program to:the execution
'she executes the program'

The constructions of E2.a and b are in a paradigmatic relationship to each other and to more elementary constructions containing a simple full verb instead of the LVC. E3 shows these more elementary counterparts to the LVCs of E2:

E3 a. das Programm wird ausgeführt
GERMAN the program gets out:carried
'the program is executed'

[5] Valency and aktionsart are the two main grammatical parameters systematizing the complex verbs studied in Schultze-Berndt 2003, §4, which again argues for a common denominator for these and LVCs.

b. sie führt das Programm aus
she carries the program out
'she executes the program'

These paradigmatic relations are schematized in S2: E3.a and b are represented in the right and left columns, respectively, of the row dubbed ‘simple full verb’, and similarly E2.a and b in the same columns of the row dubbed ‘light verb’.[6]

S2 Paradigmatic relations of German light-verb constructions

version **construction **	**active**	**passive**
simple full verb	x_{sbj} [A]$_{V.tr}$ y_{obj}	y_{sbj} wird (von x) [A]$_{V.tr.pass}$
light verb	x_{sbj} bringt y_{obj} zu [A]$_{VN.dat}$	y_{sbj} kommt/gelangt (durch x) zu [A]$_{VN.dat}$
	x_{sbj} unterzieht y_{obj} [A]$_{VN.dat}$	y_{sbj} findet/erfährt (durch x) [A]$_{VN.acc}$

In the formulas, A is a lexeme coding a DRC such as ‘execute’ in E3, VN is ‘verbal noun’. A comparison of the two columns reveals that regular German LVCs replicate the voice paradigm of the underlying simple verb. The paradigmatic relation between a simple full verb construction and a LVC (represented by the rows of S2) may be described thus: The DRC A that is coded as a simple full verb in the basic construction is coded as a non-verb, viz. a VN, in inner dependent function in an LVC. The exact syntactic function of the inner dependent varies depending on the valency of the light verb.[7]

S2 mentions the two light verbs illustrated in E2 and a couple more with very similar function. They are tabulated in T3:

T3 Productive German light verbs

active/transitive		**inactive/passive**	
form	**meaning**	**form**	**meaning**
zu [A]$_{VN.dat}$ bringen	bring	zu [A]$_{VN.dat}$ kommen	come
		zu [A]$_{VN.dat}$ gelangen	arrive
[A]$_{VN.dat}$ unterziehen	subject	[A]$_{VN.acc}$ erfahren	experience
		[A]$_{VN.acc}$ finden	find

Apart from the voice paradigm focused on here, there are also subtle shades of aktionsarten which will be foregone. It is, however, to the point to recall Germanists’ earlier verdicts on the stylistic qualities of LVCs:[8] An LVC introduces syntactic complexity into the text in first nominalizing a verb and then verbalizing the product again by means of a light verb. On the other hand, the subtle shades just alluded to are not necessarily present in each use of an LVC, and more often than not an LVC is simply synonymous with its base verb (cf. E2 and E3). In

[6] The switch in the order of a/b in E2 as opposed to left/right in S2 is due to opposite markedness relations: In the simple full verb construction (E3), passive is marked as against active, while in the LVC of E2, the light verb of the active version is a lexical causative of the light verb of the inactive version.

[7] The category index VN in the formulas is a slight simplification, since the complement shows morphological symptoms of definiteness – thus, of NP status – in prepositional forms such as *zur* ‘to:the(F)’, as in E2.

[8] There is a sizable body of literature published on German *Funktionsverbgefüge*. Engeln 1968, Helbig 1979 and Polenz 1963 may be mentioned as important early contributions.

all those cases, the added complexity is redundant. German LVCs are typical of relatively stilted written style.

The following points may be retained from this analysis:

- The regular German LVC involves a derived verbal noun as the inner dependent of the light verb.
- The LVC is, thus, in a paradigmatic relation with a simple full verb construction whose verb is precisely the base of that verbal noun.
- The formation of an LVC on the basis of a simple full verb construction involves a paradigm of light verbs. The process is highly compositional.

3.3 Light-verb constructions in Persian

According to Karimi-Doostan 1997:82, modern Persian (colloquial and standard) has about 150 simple verbs, some 30 (20 according to Family 2008:140) of which serve as light verbs.[9] Most verbal expressions involve an LVC.[10] E4 provides a few typical examples.

E4 PERSIAN
a. qadam zad-an
step beat-INF
'to go for a walk'

b. xejālat kešid-an
shame pull-INF
'to get ashamed'

c. padid ʔāmad-an
event come-INF
'to happen' (Avazeh Mache p.c.)

As is apparent, the Persian LVC has the general structure of S1 above. Members of C (the inner dependent) are called "preverbal elements" in Family 2008:140 and "non-verbal elements" in Karimi-Doostan 2006. This is an open class subsuming a heterogeneous set of word classes and syntactic categories, including N, Adj, Adv, Prep, PrepP and verb completor. The latter's members are semantically most similar to abstract nouns; many of them are Arabic loans.[11] As their precise grammatical nature is neither clear nor at stake here, the following hints may suffice: Unlike adjectives, verb completors have no degrees of comparison and do not take adverbial modifiers. Unlike nouns, they do not decline, take determiners or depend directly on a preposition. Unlike verbs, they do not conjugate or constitute a clause predicate. Some of them, however, (including, e.g., *ʔanjām* 'execution' in E7) may function as the head of an attribute linked by ezafe (AT in the gloss). For purposes of the present argument, it suffices to know that verb completors share with all the other members of C the property of being neither verbs nor derived from verbs by any regular process.

The verbs functioning as light verbs in E4 may be used as full verbs, too; but they appear far more often in LVCs. Moreover, as may be gathered from the examples, many of these

[9] The quantitative relations in the cognate Urdu are of the same order; see Butt & Geuder 2001:328.

[10] In a corpus, 2500 different LVCs were found (Karimi-Doostan 1997:83).

[11] It is Karimi-Doostan (2006) who recognizes a separate class of verb completors ("classless words"); the other authors tacitly subsume them under the class of nouns.

collocations are lexicalized and highly idiomatic. They bear no paradigmatic relation to a simple base verb construction and instead fill a lexical gap.

E5 provides a pair of examples that are in a diathetic relationship like E2. Observe that although the LVC contains an inner dependent, it can take a direct object, as demonstrated by E5.b.

E5 a. Sasan (tavasote Ali) šekast xord.
PERSIAN Sasan by Ali defeat eat(PAST)
'Sasan suffered defeat / was defeated (by Ali).'

b. Ali Sasan-rā šekast dād.
Ali Sasan-ACC defeat give(PAST)
'Ali defeated Sasan.' (Karimi-Doostan 1997:135)

Beside the prevalent type illustrated in E4f, there are a few full verbs in Persian coupled paradigmatically with a verbal noun, which latter may serve as the inner dependent of a corresponding LVC. They are enumerated in T4 and quoted as a bare stem (identical with the past tense form in most cases).

T4 Simple full verbs and verbal nouns in Persian

form / **meaning** \	**verb stem**	**verbal noun**
pay	pardāxt	pardāxt
cry	gerist	gery-e
confuse	ašoft	ašoft-e
moan	nalid	nal-e
strive	kušid	kuš-eš
choose	pasandid	pasand
live	zist	zendagi

Each of the verbal nouns of the right-hand column in T4 combines as an inner dependent with the light verb *kardan* 'do' to form an LVC that is essentially synonymous with the simple verb,[12] as illustrated in E6.

E6 a. man pul-rā pardāxt-am
PERSIAN I money-ACC pay(PAST)-1.SG
'I paid the money'

b. man pul-rā pardāxt kard-am
I money-ACC pay do(PAST)-1.SG
'I paid the money' (Karimi-Doostan 1997:62)

In contrast with the German system, however, this kind of paradigmatic relation is not constitutive of the Persian LVC. There is no semantic or structural unity to the set of T4. Its entries have, in fact, been arranged in such a way as to bring out what regularity there is; but as may be seen, hardly any two entries follow the same structural pattern. The only regular and productive way of forming a verbal noun is the infinitive (illustrated by E4), which may

[12] There are two more such pairs, *qalt-id* 'roll' (intr.) - *qalt xord* (roll eat) and *farift* 'trick, fool' - *farib dād* (trick give), involving different light verbs.

be formed for any verb, including those in T4. The infinitive, however, cannot occupy the position of C in S1. There is, consequently, no regular process of deriving an LVC from a simple full verb. Instead, Persian LVCs are organized internally in rather tight paradigms formed by the parameters of voice and aktionsart. Here only voice will be noted. A first example of a diathetic opposition between inactive or passive (E5.a) vs. active and transitive (E5.b) was already seen. Similarly in E7:

E7 a. kār-am ʔanjām šod / yāft
PERSIAN business-POSS.1.SG execution become(PAST) / find(PAST)
'my business got executed' (Karimi-Doostan 1997:113/129)

b. Ali ʔin kār-rā ʔanjām dād
Ali this business-ACC execution give(PAST)
'Ali executed this job' (Karimi-Doostan 1997:79)

This paradigmatic relation is schematized in S3, which accounts for some of the light verbs illustrated so far.

S3 Diathetic relation of Persian light-verb constructions

active/transitive	**inactive/passive**
x_{sbj} y_{obj} [[A]$_C$ dād / kard]$_{LVC}$	y_{sbj} (tavasote x) [[A]$_C$ šod / xord]$_{LVC}$

In the formulas, A represents a DRC, C is any of the categories mentioned before, including the verb completor. T5 contains the core paradigm of light verbs (cf. Karimi-Doostan 1997:83f), some of which produce over 500 LVCs (Family 2008:146):

T5 Productive Persian light verbs

active/transitive		**inactive/passive**	
form	**meaning**	**form**	**meaning**
kard	do	šod	become
dād	give	xord	eat, undergo
zad	hit	yāft	find
ʔāvard	bring	ʔāmad	come

The only productive process of forming verbal nouns is the infinitive; and it also serves in the nominalization of LVCs, as pointed up by the bracketing in E8f:

E8 hoquq-e kam baʔes-e rešve xord-an ziyad šod-e ast
PERSIAN salary-AT small cause-AT [bribe eat-INF much] become-PTCP is
'low salaries have become the cause of much bribery' (Family 2008:157)

E9 Sohrāb as dast dād-an-e Rostam xeyli afsus xord.
PERSIAN Sohrab [from hand give-INF-AT Rostam] much regret eat(PAST)
'Sohrab heavily regretted the loss of Rostam.' (Family 2008:149)

The following generalizations about the Persian light-verb system may be retained:

- There are only about 150 verbs, at most 30 of which may function as light verbs.
- Most DRCs are not lexicalized as verbs, but as nouns, adjectives, adverbs and verb completors. In order to function as predicates, these combine with a light verb.

- LVCs are highly productive, both in their semantically regular and in their lexicalized variant. They represent, in fact, the most important process of verb formation.
- Most LVCs lack a more basic counterpart formed by a simple full verb.

3.4 Comparison

We have seen LVCs in German and Persian alike. However, their locus in the language system is essentially different. One of the basic differences concerns the categorial status of the words that may function as the inner dependent of an LVC (A in S1). T6 opposes some representative examples.

T6 DRCs in German and Persian

language / concept	German	Persian
performance	[[Ausführ-]$_{V}$ -ung]$_{VN}$	ʔanjām$_{VC}$
continuation	[[Fortsetz-]$_{V}$ -ung]$_{VN}$	ʔedāme$_{VC}$
consideration	[Rücksicht]$_{VN}$	mahsub$_{VC}$
persuasion	[[Überred-]$_{V}$ -ung]$_{VN}$	vādār$_{VC}$

As T6 is meant to illustrate, the German abstract nouns are mostly derived by regular nominalization from a verb stem (even *Rücksicht* is so derived, although not in a regular way).[13] Again, all of the Persian entries are verb completors (the first two possess at least one nominal property, viz. they take an attribute; Karimi-Doostan 2006). They bear no derivational relationship to verbs. On the other hand, while all of the Persian entries have their *raison d'être* as inner dependent of an LVC, half of the German abstract nouns of T6 (and most abstract nouns that do not appear in T6) are seldom or never used in LVCs.

As most of the Persian LVCs do not have a simpler counterpart, the status of the LVC in the syntactic system differs in the two languages, too. While in German, an LVC is clearly a complex construction, formally, semantically and stylistically marked against its base, the Persian LVC is in most cases the simplest construction available; it is one of the basic verbal constructions.[14]

These interrelated differences between the German and Persian LVC are summarized in T7.

[13] Given that the verbal base of *Rücksicht* is no longer available, this abstract noun may be seen as moving towards the pole of a basic verb completor used primarily in LVCs, with *nehmen* as the pertinent light verb (*Rücksicht nehmen* 'consider'). As a consequence, the nominalization of the LVC, viz. *Rücksichtnahme*, increasingly replaces the basic nominalization *Rücksicht*.

[14] This finding is strikingly similar to the one reported in Butt & Geuder 2001:358 concerning two languages cognate with German and Persian, resp.: "While in English the employment of light verbs appears to have more the status of a primarily stylistic device, the use of light verbs is very entrenched in Urdu."

T7 LVCs in German and Persian

language \ criterion	German	Persian
lexical category of dynamic relational concepts	mostly verbs	mostly non-verbs
lexical basis of regular LVC	simple full verb which is nominalized	basic non-verb
complexity status of LVC	marked	basic

Finally, there is nominalization in both languages. However, in German it operates at a lower level of structural complexity, as it applies to full verb stems and derives verbal nouns from them. In Persian, again, most abstract lexemes are not transparently derived from verbs. It is only at the level of the verb group and the clause that the infinitive produces a nominalized clause. And it applies to LVCs alike, producing nominalizations based on LVCs rather than the other way around, as in German.

In a dynamic perspective, this means that both languages have the means to derive nominal from verbal expressions and vice versa. However, since they opt for opposite bases, operations that produce constructions of higher-level complexity are cross-linguistically out of phase. S4, read from bottom to top, is meant to illustrate this by showing levels of structural complexity in horizontal rows and marking cross-linguistically like syntactic categories with the same color.

S4 Categorization of DRCs in German and Persian

level / language	German	Persian
complex 2	[[A]$_{VN}$ [B]$_{LV}$]$_{LVC=V}$	[[A [B]$_{LV}$]$_{LVC}$ -INF]$_{VN}$
↑		
complex 1	[A-NMZR]$_{VN}$	[A [B]$_{LV}$]$_{LVC=V}$
↑		
basic	[A]$_{V}$	[A]$_{VC}$

- German (left-hand column of S4) opts for categorizing most DRCs (A) in one major word class, viz. the verb. There is, accordingly, nominalization at the stem level, i.e. a derivational operation that creates verbal nouns. These, in turn, may form the lexical basis of another process of verbalization, producing LVCs which double simple full verb constructions at a higher level of complexity.
- Persian (right-hand column) opts for categorizing most DRCs in a set of non-verbal word classes, some of which correspond semantically to *nomina actionis* of other languages and which crystallize in the verb completor used as category index in the base row of S4. LVCs are at the next higher level of complexity, and they are produced by a one-step operation of verbalization (combining the non-verb with a light verb); apart from the exceptions illustrated in T4, they double nothing. LVCs, in turn, may be nominalized at

the syntactic level, producing infinitivals which fall into the same category as a subset of the basic abstract lexemes.

It should be conceded that the picture is somewhat idealized. By the lexicalized collocations exemplified in E1, German deviates from the pole of a language that only uses derived *nomina actionis* in LVCs. Likewise, Persian deviates from the pole of a language having only LVCs at the base level by its 100 or so simple full verbs which do not serve as light verbs. Probably a pair of languages can be found to instantiate the extreme poles more clearly. Moreover, it must be born in mind that even a language relying on verb completors in a more principled way than Persian does possess verbs, so that the contrast analyzed here cannot be absolute in principle.

At the same time, that is a continuum, not a clean complementary distribution of two properties across languages. There are languages that make extensive use of both of the converse strategies of S4. Korean, for instance, has both (predominantly native) simple full verbs nominalized by the left-hand strategy of S4 and verb completors (predominantly Chinese loans) combined with a light verb (*hata* 'do') to yield a basic LVC by the right-hand model of S4. In this field as in other areas of typology, the primary contrast is not between languages, but between constructions and strategies. Languages instantiate (contrasting) types only to the extent that they cling exclusively to one of the available structural models.

4 Some further cases

4.1 Base transitivity: basic transitivity vs. intransitivity

Certain DRCs like 'break' and 'burn' essentially imply an undergoer and are compatible both with the presence and with the absence of an actor. With an allusion to Chafe 1970, ch. 11, these will here be called action-processes. There are two opposite options in coding the distinction between presence vs. absence of an actor by regular derivation (cf. Haspelmath 1993, Nichols et al. 2004, Koptjevskaja-Tamm 2008:32): An action-process may be coded basically as an intransitive verb, which may be causativized if an actor is involved; or alternatively, it may be coded basically as a transitive verb, which may be anticausativized if no actor is involved. Apart from these two strategies, there are other solutions to the problem. One is to leave the distinction uncoded, i.e. to operate with labile verbs. Another is to shift the problem into the lexical sphere: thus, there may be equally elementary or equally derived lexemes for both the intransitive and the transitive version, so that none is based on the other.

Anyway, some languages adhere rather consistently to one or the other of the two opposite principled solutions. Like Bororo, Coast Salish and several other Amerindian languages, Japanese categorizes action-processes preferably as intransitive verbs, while Russian categorizes them preferably as transitive verbs. Japanese has causativization, but no morphological anticausativization. In Russian, it is the other way around: it has anticausativization (in the form of reflexive verbs), but no morphological causativization. For the concepts illustrated in T8, the basic verbs in Japanese mean 'move' (itr.), 'get a fright' and 'get angry', while the basic Russian verbs mean 'move' (tr.), 'frighten' and 'annoy'.

T8 Basic intransitivity and transitivity in Japanese and Russian

language / verb class concept ↘	Japanese		Russian	
	intransitive →	transitive	intransitive ←	transitive
move	ugok-u	ugok-ase-ru	dvigat'-sja	dvigat'
scare	odorok-u	odorok-as(e-r)u	pugat'-sja	pugat'
annoy	okor-u	okor-ase-ru	serdit'-sja	serdit'

S5, again to be read from bottom to top, visualizes the converse relationship between the base and derived constructions in the two languages.

S5 Categorization of transitive agentivity in Japanese and Russian

derived	[X-CAUS]$_{V.tr}$	[X-ANTICAUS]$_{V.itr}$
↑		
basic	[X]$_{V.itr}$	[X]$_{V.tr}$
level / language	**Japanese**	**Russian**

Haspelmath 1993 and Nichols et al. 2004 make it clear that no language in their sample adheres exclusively to one of the alternate basic categorizations.[15] Moreover, if alternative p is chosen as basic, this does not entail that alternative q has to be derived from p. Thus, this relationship between base transitivity and base intransitivity is, again, not really a clean case of complementary distribution of strategies across languages. It may, however, be retained that if a language makes a principled decision for one basic categorization (i.e. it employs uniform basic categorization as explicated in §5), then it needs an operation of recategorization, whereas if it does not, then it does not need such recategorization operations, either. For instance, English has many labile verbs such as *move*, *boil* and *break*; and it has neither a morphological operation of causativization nor one of anticausativization.

4.2 Spatial regions: adpositions and relational nouns

A **spatial region** is an aspect of the topological structure associated with a physical object by virtue of its being three-dimensional and occupying a position in three-dimensional space. Examples include ‘top’, ‘front’, ‘interior’, ‘vicinity’. All languages code them in some way, since they are an integral part of human orientation in space, more specifically, of the “intrinsic frame of reference” (Levinson & Wilkins 2006:3f). However, their assignment to some part of speech is not straightforward. It depends less on their conceptual nature and more on the way they are used in syntactic constructions. Such concepts are used primarily as S in propositions of the general form of S6[16] as illustrated by E10:

[15] According to the description in Crowell 1979, Bororo (Macro-Gê) instantiates the base-intransitive pole of the continuum. According to Storch 2009:124, Hone (Jukun) instantiates the opposite pole, i.e. a language without intransitive verbs.

[16] Cf. Lehmann 1992. For a complete theoretical basis, manner of motion would be required as a further semantic component, as in Talmy 1985 and Levinson & Wilkins 2007. ‘Trajector’ and ‘reference object’ correspond to the

S6 Semantic structure of local situation

Trajector T rests/moves in local relation L to spatial region S of reference object O.

E10 Linda went into the house.

In E10, T = Linda, L = allative, S = interior, O = house. As S6 makes clear, spatial regions are conceptually relational, as they are a topological property of a first-order entity (O). If that conceptual property is reflected in grammar, O is commonly coded as a complement governed by S. That is the case if S appears as a relational noun like 'top', 'interior', as illustrated by E11a.

E11 a. A journey to the interior of the earth (Jules Verne)
b. Linda looked at the interior.

Nominal government is not very distinctive in English, as sentences like E11b (lacking a complement) are freely admissible. In such cases, the semantic argument place for O is filled deictically or anaphorically.

The chief alternative to the relational noun in the categorization of spatial regions is the adposition, as in E12.

E12 a. Linda went inside the house.
b. Linda went inside.

Adpositions like English *inside* (*within*, *above*, *below* …) may be described as having an optional complement. Already Jespersen (1924:88f) likened the difference between the adposition (as in E12a) and the adverb (as in E12b) to the difference between a transitive and an intransitive verb, arguing that it was a minor distinction. In the following, we will use the terms 'adposition' and 'relational noun' with the understanding that they inherit the semantic relationality of the concept of the spatial region but that their government of a complement may be pronounced to different degrees; i.e. we will ignore the difference between an adposition and an adverb.

The two strategies of categorizing the spatial region either as a relational noun or as an adposition thus do not differ essentially in their treatment of the relation of S to O. Instead, they differ in their treatment of the relation of S to the situation core. The categorization of the spatial region as adposition presupposes its use as modifying the situation core, while its categorization as a relational noun implies nothing in that respect. It might appear that the latter choice is more economic, since the way that the spatial region word depends on its syntactic head may actually vary. In E11, it is, in fact, governed (through the preposition governed by the verb), so that no modification is involved. However, given their lack of ontological autonomy, spatial regions seldom constitute referential objects, which would require them to be represented by an NP, as in E11b. In most cases, S bears some local relation L to the situation core, as assumed in S6; and that may perfectly well be provided by an adposition, as it is in E12.

The principal local relations are essive (rest) vs. lative (motion), the latter subdividing into allative, ablative or perlative. The particular local relation is an independent component of S6

latter's 'figure' and 'ground', resp. Their conception is, however, at variance with the one proposed here in that it assumes (p. 3) topology to be relevant only in "stasis" (i.e. rest) as opposed to motion.

and may be coded independently, as proved by the substitutability of *to* (allative) by *through* (perlative) in E11a. In principle, the adposition might imply a semantically empty modifying relation to its head (simply speaking, the verb of motion or rest); and the precise nature of L would be coded independently. This comes down to accompanying the spatial region adposition by a local relator (adposition or case affix). Morphological microanalysis proves that to obtain in E10: *into the house* is [*in* [*to* [*the house*]]], where *in* is the adpositional adverb representing the spatial region, while *to* is the local relator specifying the allative relation. Alternatively, the local relation may be part of the meaning of the head, as it is in English *approach* (allative) and *leave* (ablative). In the language to be considered below, German, spatial region and local relation are not coded separately in a regular way, since some of the adpositions coding spatial regions are restricted to a subset of the local relations.

Since spatial regions are so intimately bound up with spatial orientation, they are often lexically merged with other semantic components of a spatial situation, not only with L, but also with the concept of motion itself. That is the case, for instance, for a verb meaning 'enter', which codes an allative local relation to the interior spatial region of the reference object. In order to get a clear alternative, we will here concentrate on the lexicalization of spatial regions separate from verbs. That then reduces to the alternative already introduced, relational noun vs. adposition.

Japanese is among the languages which primarily categorize spatial regions as relational nouns. These include nouns such as *yoko* 'side', *ue* 'top', *usiro* 'back' and the like (see T9 below for a fuller enumeration). They take a genitive complement, as in *heya-no naka* (room-GEN inside) 'interior of the room'. Again, local relators are coded as cases (agglutinative suffixes or enclitic postpositions depending on the analysis), viz. *-ni* ALLATIVE-LOCATIVE, *-de* PERLATIVE, *-kara* ABLATIVE. These attach to nouns, including relational nouns of spatial region, and together with the latter produce complex postpositions, as in (*heya)-no naka-kara* (room-GEN inside-ABL) 'out of (the room)'. Finally, regional property concepts such as 'upper' are coded as genitive attributes, as in *ue-no (hako)* (top-GEN box) 'upper (box)'.

German, on the other hand, opts for the alternative of primarily categorizing spatial regions as adpositions (including adverbs). These are prepositions such as *über* 'above', *unter* 'under', *vor* 'in front of' and adverbs such as *oben* 'above', *unten* 'below' and *vorn* 'in front'. Thus, in contradistinction to Japanese, spatial relators modifying their head are elementary forms. As may be seen, some of the prepositions are related diachronically to corresponding adverbs, sharing certain submorphemic material with them, so it may be possible to reconstruct underlying relational nouns there. Anyway, in Modern German there are no primitive nouns of spatial regions, with one exception, *Seite* 'side'.[17] Instead, such nouns have to be formed in a cumbersome way. First, a region adjective is derived from the preposition/adverb in a partly regular fashion,[18] yielding adjectives like *ober(-e*), 'upper', *unter(-e)* 'lower', *vorder(-e)* 'front'. Their stems then combine as attributes with the generic region noun *Seite* to form compound nouns such as *Oberseite* 'top', *Unterseite* 'bottom', *Vorderseite* 'front'. At this point, finally, the semantic equivalent to the Japanese basic region noun is reached.

[17] There are two more primitive nouns in this lexical field, *Grund* 'bottom' and *Spitze* 'top', which however lack the meanings relevant to the paradigm of T9 and instantiated by *Unterseite* and *Oberseite*, resp.

[18] The details are irregular. In principle, the adjective is based on the root just as the adverb and adposition are. Two of the compound nouns do not even contain an adjective and instead the adverb.

As may be seen, the two languages form their basic expressions at opposite points and employ converse morphological and syntactic operations to reach the point that the other language started from. This is visualized in T9 for the basic spatial regions.

T9 Categorization of spatial regions in Japanese and German

language / **word class** / **region** ＼	**Japanese**			**German**		
	noun	→ **adverb/adposition**		**noun**	← **adverb/adposition**
side	yoko	(NP_{gen}) yoko-ni		Seite	neben NP_{dat}
top	ue	(NP_{gen}) ue-ni		Oberseite	oben/auf (NP_{dat})
bottom	sita	(NP_{gen}) sita-ni		Unterseite	unten/unter (NP_{dat})
front	mae	(NP_{gen}) mae-ni		Vorderseite	vorn/vor (NP_{dat})
back	usiro	(NP_{gen}) usiro-ni		Hinterseite	hinten/hinter (NP_{dat})
interior	naka	(NP_{gen}) naka-ni		Innenseite	innen/in (NP_{dat})
exterior	soto	(NP_{gen}) soto-ni		Außenseite	außen/aus (NP_{dat})
right side	migi	(NP_{gen}) migi-ni		rechte Seite	rechts (NP_{gen})

S7 is again to be read from bottom to top and visualizes the complementary categorization of spatial regions in Japanese and German and their converse operations of creating constructions that are basic in the respective other language.

S7 Converse categorization of spatial regions in Japanese and German

derived	[NP-GEN **X-CASE**]$_{AdpP}$		[**X(-ADJVZR)-*Seite***]$_{N.rel}$
↑			
basic	[**X**]$_{N.rel}$		[[**X**]$_{Adp}$ NP-CASE]$_{AdpP}$
level ／ **language**	**Japanese**		**German**

This case study, once more, does not show a complementary distribution of two features across a sample of languages, but instead an alternative choice made by two languages in some well-defined domain, each with opposite consequences for adjacent parts of the linguistic system.

4.3 Other examples

Many more areas of grammar could be analyzed in the way proposed here for light-verb constructions, base transitivity and spatial regions. The following four areas have partly been studied before, but would be worth taking up in this perspective:

- Properties: Apart from the primary categorization of property concepts as abstract nouns (like 'beauty'), resorted to by relatively few languages, the main alternative is between categorizing them as adjectives ('beautiful') or as stative verbs ('be beautiful'). There are then operations of transferring property words into the respective other category. See Lehmann 1990, §4.
- Nomination: For the task of designating entities of high time-stability ("things"), languages have the nominal category. Given this, there is still the alternative of treating it as basic, i.e. providing a large stock of elementary nouns, or forming most nouns by nominalization of verbs. The alternatives are called 'labeling vs. descriptive' in Seiler 1975. The case is complementary to the one of light-verb constructions treated in §3; here the categorization of static non-relational concepts is at stake, there the categorization of DRCs was treated.
- Kinship concepts may be categorized either as nouns (as in English and Yucatec Maya) or as verbs. Again, in the former case, they may undergo a verbalization operation (like Yuc. *atan-t-* 'have as one's wife'), while there may be a nominalization in the latter case (as in Yuma, Halpern 1942).
- Valency and serialization: Their partly complementary relationship may be conceived in terms of two types. A language of type 1 distinguishes several adverbal complement and adjunct relations and allows nominal dependents bearing such relations to cluster syntagmatically on one verb. Its verbs may be up to quadrivalent. It has no verb serialization. German is a case in point. A language of type 2 makes no formal distinction between different types of verbal dependents. Its highest verb valency is bivalent. It uses serialization in order to combine more participants into one situation. !Xun (Khoisan) is an example (König 2009:23).

5 Explicating the hypothesis

The general thesis developed here on the basis of a specimen analysis and further evidence evoked more summarily will now be made as general and explicit as possible. Assume first a set of interlingual syntactic categories that apply to linguistic signs both at the root/stem level and at higher complexity levels such as the syntagm and the clause, such as absolute and relational nominal, transitive and intransitive verbal, absolute and relational adjectival, adverbial, adpositional etc. (word form or phrase). Assume second a particular set of lexical concepts[19] which belong to one conceptual field or domain D but are usable in two different syntactic categories C_i and C_j.

Next, the thesis presupposes the notion of **uniform basic categorization** of such a conceptual domain. Assume a set of syntactic categories of the kind mentioned. The basic categorization

[19] Just like the term *verbal concept* evoked at the beginning of §4.1, the term *lexical concept* is current, but not particularly ingenious since the fact that the concepts in question tend to be coded in lexemes is an observation *a posteriori*. The term is responsible for the awkwardness of such formulations as of lexical concepts that are not lexicalized (further below). Expressions like *specific concepts* would be more appropriate, but would require an explanation on every occasion of their use.

of a concept is its coding at the lowest level of structural complexity which is at all categorized in the language.[20] Uniform basic categorization of some conceptual domain is basic categorization of all of its members in the same category. Naturally, conditions for uniform categorization are optimal if D is semantically homogeneous. For instance, the categorization of action-processes as transitive or intransitive verbs may be sensitive to whether the DRCs in question imply a specific instrument, which in turn calls for an agent (Haspelmath 1993). Thus, 'cut' will preferably be categorized as a transitive verb even in languages that otherwise prefer basic intransitivity because it implies a specific instrument. Such semantic differences are the main factor responsible for multiplex categorization of a set of concepts. On the other hand, such a semantic difference does not necessarily exclude the lumping of all the concepts concerned into one structural category.

Now given such a conceptual domain D and two alternative categories C_i and C_j that it fits into. Then the general hypothesis suggested by the data analyzed here may be phrased as follows:

> If a language assigns D uniformly to C_i at the basic level, it will possess an operation $O^{\rightarrow Cj}$ of converting members of D into C_j at some higher level of grammatical complexity; and symmetrically, if a language assigns D uniformly to C_j at the basic level, it will possess an operation $O^{\rightarrow Ci}$ of converting members of D into C_i at a higher level.

Thus, the two couplings of basic categorization plus recategorization operation form mirror images. It may be profitable to develop a quantitative version of the hypothesis by quantifying the extent to which members of D are assigned to a certain category instead of categorically requiring uniform basic categorization. As it stands, the binary contrast suggested by the hypothesis depends on the condition of uniform basic categorization. It should be clear that even if there are just two categories in which a given D is used, there is no necessity for a language to make a principled choice for its basic categorization; there are common ways of eschewing the decision. One solution is category indeterminacy: The concepts in question may simultaneously belong to categories C_i and C_j. Thus, failing to decide the alternative of categorizing DRCs as either verbs or non-verbs results in a lexeme class that is indeterminate between these two categories, like English *walk* and *cry*, which may be used as either nouns or verbs. For the alternative between transitive and intransitive verbs, category indeterminacy yields labile verbs, again a solution favored by English. For the alternative of categorizing spatial regions as either relational nouns or as adpositions/adverbs, English once more provides examples of lexemes like *back*, *left* and *right*, which are used in either category. Another way of eschewing the alternative is to provide two lexemes for each of the concepts in question, one in each category. In the field of DRCs, that is the case of English *choose* vs. *choice*, *live* vs. *life*. For action-processes, English has lexical pairs such as (*be) angry* vs. *annoy*, *eat* vs. *feed*. Syntactic doublets of spatial regions include *on* and *top*, *under* and *bottom*. In all these fields, there is no necessity for a language to make a principled choice. Only if it does will it need an operation to make the respective other categorization available.

Under these provisos, the hypothesis formulated above is to be taken as an empirical hypothesis. It can be falsified as follows: Provide a set of interlingual syntactic categories of

[20] The lowest level of categorization is the root for some languages, the stem (or even the phrase) for others; s. Lehmann 2008.

the kind mentioned, such that every language has a subset of them. Identify a conceptual domain D whose elements are coded by lexemes in your sample of languages. Determine, for each language L of the sample, in which category it primarily codes D. (If all the languages of the sample assign D to the same category, try another domain or enlarge the sample. For instance, punctual event concepts are seldom categorized as anything but verbs at the basic level.) If D is assigned to more than one category C_i, C_j ..., find two languages L_1 and L_2 that fulfill the following condition: Both L_1 and L_2 possess both C_i and C_j, but L_1 has uniform basic categorization of D in C_i, while L_2 uniformly assigns D to C_j. Finally, show that there is no way in L_1 of transferring elements of D into C_j, and conversely for L_2. With the last move, you have falsified the hypothesis. Obviously, the hypothesis will be the easier to falsify the narrower the concept of 'transferring an element into a category' is defined.

6 Conclusion

Why should a typology of converse categorization strategies be important for our understanding of the workings of language?

- Both the basic category assignment and the presence of the associated recategorization operation shape the structure of sentences and of texts in the language. For instance, the stylistic role of LVCs in Persian and German may be assessed more objectively once their typological bases are clarified.
- For a language to provide uniform basic categorization for some conceptual domain coupled with an operation of recategorization at the morphological or syntactic level means for it to either introduce relatively more grammatical structure into its lexicon (cf. Coseriu's [1976, §5.2] idea of word-formation as a "grammaticalization of the lexicon") or to burden its syntax with categorization operations. It remains to find out what the typological principles underlying these options are, under what conditions the balance between grammar and lexicon may be tilted in either direction, and if the option of converse categorization strategies is chosen, what conditions either of the converse choices.
- Cases of converse categorization provide insight into the role of categorization in language and make us see the degrees of freedom obtaining there. These concern both the individual categories with their relations to each other and the levels of structural complexity at which concepts are categorized.

Thus, analyzing such distributions of coding strategies across languages may be a powerful heuristic in systematizing cross-linguistic variation and finding out about the principles underlying the alternative of solving certain tasks of cognition and communication either in the lexicon or in the grammar.

References

Benveniste, Émile 1957, "La phrase relative, problème de syntaxe générale". *BSL* 53:39-53.

Butt, Miriam & Geuder, Wilhelm 2001, "On the (semi)lexical status of light verbs." Riemsdijk, Henk C. van & Corver, Norbert (eds.), *Semi-lexical categories. The function of content words and the content of function words*. Berlin & New York: Mouton de Gruyter (Studies in Generative Grammar, 59); 323-370.

Chafe, Wallace L. 1970, *Meaning and the structure of language*. Chicago & London: Chicago University Press.

Coseriu, Eugenio 1976, "Die funktionelle Betrachtung des Wortschatzes." *Sprache der Gegenwart* 14:7-25.

Crowell, Thomas H. 1979, *A grammar of Bororo*. Cornell University PhD diss. Ann Arbor & London: University Microfilms International.

DeLancey, Scott 2005, "Adpositions as a non-universal category." Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.), *Linguistic diversity and language theories*. Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 72); 185-202.

Dryer, Matthew 2005, "Order of adposition and noun phrase." Haspelmath et al. (eds.) 2005, ch. 85.

Engeln, Bernhard 1968, "Zum System der Funktionsverbgefüge." *Wirkendes Wort* 18:289-303.

Family, Neiloufar 2008, "Mapping semantic spaces. A constructionist account of the «light verb» *xordæn* «eat» in Persian." Vanhove, Martine (ed.), *From polysemy to semantic change. Towards a typology of lexical semantic associations*. Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 106); 139-161.

Greenberg, Joseph H. 1963, „Some universals of grammar with particular reference to the order of meaningful elements.“ Greenberg, Joseph H. (ed.), *Universals of language. Report of a conference held at Dobbs Ferry, New York, April 13-15, 1961*. Cambridge, MA: MIT Press; 58-90.

Halpern, Abraham M. 1942, “Yuma kinship terms”. *American Anthropologist* 44:425-441.

Haspelmath, Martin 1993, "More on the typology of inchoative/causative verb alternations." Comrie, Bernard & Polinsky, Maria (eds.), *Causatives and transitivity*. Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 23); 87-120.

Haspelmath, Martin & Dryer, Matthew & Gil, David & Comrie, Bernard (eds.) 2005, *The world atlas of language structures*. Oxford etc.: Oxford University Press.

Helbig, Gerhard 1979, "Probleme der Beschreibung von Funktionsverben im Deutschen." *Deutsch als Fremdsprache* 16:273-285.

Jespersen, Otto 1924, *The philosophy of grammar*. London: G. Allen & Unwin; New York: H. Holt & Co.

Karimi Doostan, Gholamhossein 1997, *Light verb constructions in Persian*. PhD dissertation. Essex: Essex University.

Karimi-Doostan, Gholamhossein 2006, “Classless words in Persian”. Paper presented at the Conference ‘Universality and Particularity in Parts of Speech Systems’, Amsterdam, 8-9 June, 2006. [Abstract: http://home.hum.uva.nl/pos/abstracts/Abstract_Doostan.pdf; consulted 25/11/2009]

König, Christa 2009, "!Xun." Dimmendaal, Gerrit J. (ed.), *Coding participant marking. Construction types in twelve African languages.* Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 110); 23-53.

Koptjevskaja-Tamm, Maria 2008, "Approaching lexical typology." Vanhove, Martine (ed.), *From polysemy to semantic change. Towards a typology of lexical semantic associations*. Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 106); 3-52.

Lehmann, Christian 1990, "Towards lexical typology." Croft, William & Denning, Keith & Kemmer, Suzanne E. (eds.), *Studies in typology and diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday.* Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 20); 161-185.

Lehmann, Christian 1991, "Grammaticalization and related changes in contemporary German". Traugott, Elizabeth C. & Heine, Bernd (eds.), *Approaches to grammaticalization*. Vol. II: *Focus on types of grammatical markers*. Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 19:2); 493-535.

Lehmann, Christian 1992, "Yukatekische lokale Relatoren in typologischer Perspektive." *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 45:626-641.

Lehmann, Christian 2002, *Possession in Yucatec Maya.* Erfurt: Seminar für Sprachwissenschaft der Universität (ASSidUE, 10) (2nd revised edition).

Lehmann, Christian 2008, "Roots, stems and word classes." *Studies in Language* 32:546-567.

Levinson, Stephen C. & Wilkins, David P. 2006, "Background to the study of the language of space." Levinson, Stephen C. & Wilkins, David P. (eds.), *Grammars of space. Explorations in cognitive diversity*. Cambridge: Cambridge University Press; 1-23.

McGregor, William B. 2002, *Verb classification in Australian languages*. Berlin: Mouton de Gruyter.

Nichols, Johanna & Peterson, David A. & Barnes, Jonathan 2004, "Transitivizing and detransitivizing languages." *Linguistic Typology* 8:149-211.

Pawley, Andrew & Lane, Jonathan 1998, "From event sequence to grammar. Serial verb constructions in Kalam." Siewierska, Anna & Song, Jae Jung (eds.), *Case, typology and grammar. In honor of Barry J. Blake*. Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 38); 201-227.

Polenz, Peter von 1963, *Funktionsverben im heutigen Deutsch: Sprache in der rationalisierten Welt.* Düsseldorf: Schwann (Wirkendes Wort, Beiheft, 5).

Sakel, Jeanette 2007, "The verbness markers of Mosetén complex predicates." Miestamo, Matti & Wälchli, Bernhard (eds.), *New challenges in typology. Broadening the horizons and redefining the foundations*. Berlin & New York: Mouton de Gruyter (Trends in Linguistics, 189); 314-335.

Schultze-Berndt, Eva 2003, "Preverbs as an open word class in Northern Australian languages: synchronic and diachronic correlates." *Yearbook of Morphology* 2003:145-177.

Seiler, Hansjakob 1972, "Universals of language." *Leuvense Bijdragen* 61:371-393.

Seiler, Hansjakob 1975, "Die Prinzipien der deskriptiven und der etikettierenden Benennung." Seiler, Hansjakob (ed.), *Linguistic workshop III. Arbeiten des Kölner Universalienprojekts 1974*. München: Fink (Structura, 9); 2-57.

Seiler, Hansjakob 1998, "On complementarity." Janse, Mark & Verlinden, An (eds.), *Productivity and creativity. Studies in general and descriptive linguistics in honour of E.M. Uhlenbeck*. Berlin & New York: Mouton de Gruyter (Trends in Linguistics, 116); 231-242.

Seiler, Hansjakob 2010, "Le continuum linguistique et la relation entre un invariant et ses variantes." *Bulletin de la Société de Linguistique de Paris* 105(1):7-33.

Storch, Anne 2009, "Hone." Dimmendaal, Gerrit J. (ed.), *Coding participant marking. Construction types in twelve African languages.* Amsterdam & Philadelphia: J. Benjamins (Studies in Language Companion Series, 110); 123-140.

Talmy, Leonard 1985, "Lexicalization patterns: semantic structure in lexical forms." Shopen, Timothy (ed.), *Language typology and syntactic description. Vol. III: Grammatical categories and the lexicon.* Cambridge etc.: Cambridge University Press; 57-149.