

3 Comparing verbal valency across languages

1 Introduction

That different verbs may have different valencies even when they are semantically similar has long been well known (at least since Tesnière 1959), and it is the reason why dictionaries contain (or should contain) valency information for each verb. For example, we need to specify that English *wait* takes a *for*-complement, while *expect* takes a direct object.

That “the same” verbs across languages also often differ in valency is not quite so well known, though typological-comparative and contrastive works in linguistics have often discussed valency mismatches of the type illustrated in (1)–(2).

(1) a. English

I₁ miss you₂.

b. French

Vous₂ me₁ manquez.

‘I miss you.’ (Literally something like ‘You are missing to me.’)

(Tesnière 1959: § 123.2)

(2) a. English

She₁ filled the container₂ with water₃.

b. Chintang (Tibeto-Burman, Nepal)

Huĩsa-ŋa₁ cuwa-Ø₃ gagri-be₂ phatt-e.

she-ERG water-NOM container-LOC fill-PST

‘She filled the container with water.’

(Literally ‘She filled the water into the container.’)

(Bickel et al. 2010: 387)

Tesnière called such cross-linguistic mismatches “metataxis” (Tesnière 1959; cf. Koch 1994), and it is such differences between languages that we are primarily interested in here.

But what does it mean for the valencies of two verbs to fail to match? How can we compare valencies across languages? These are not trivial questions, but answers to them are a prerequisite for any comparative or contrastive research on valency. This chapter will explain how we approached this problem, and what decisions we took for the Valency Patterns Leipzig (ValPaL) database (Hartmann et al. 2013).

The basic principle is that we decided to match verbs across languages on the basis of verb meaning, and to match arguments across languages on the basis of

individual argument meaning, i.e. argument microroles (see §5, and also Hartmann et al. 2014).

2 Valency: coding frames and role frames

The **valency** of a verb is the range of syntactic properties of other elements of the clause that depend on the particular choice of verb, i.e. that are verb-specific. These other elements of the clause are called **arguments**. The most salient argument properties are the **coding properties**, i.e. flagging (case or adpositional marking) and indexing (i.e. bound person marking associated with the verb).¹ Examples of argument-coding elements are Nominative and Accusative case and the preposition *of* in (3) from English, and Dative and Nominative case as well as 3rd person singular Subject indexing in (4) from German.

(3) English

[They]_{NOM} *accused* [her]_{ACC} [*of plagiarism*].

(4) German

[*Den Kindern*] *gefällt* [*der Schneemann*].
 the.PL.DAT child.PL.DAT please.3SG the.SG.NOM snowman.SG.NOM
 ‘The children like the snowman.’

Everything else in the clause is independent of the verb: A clause can contain locational and temporal setting adverbials (e.g. 5), or it can contain manner adverbials or illocutionary adverbials (e.g. 6), regardless of the kind of verb that is chosen. Such verb-independent elements are called **adjuncts**.

(5) *Last week they accused her of plagiarism in her school.*

(6) *In all frankness, they accused her with verve.*

A clause can contain subordinators or diverse particles, it can exhibit special word order, and it can occur in different tenses (all illustrated by (7) from German), independently of its verb.

(7) *weil der Schneemann dem Kind doch gefiel*
 because the snowman.SG.NOM the child.SG.DAT after.all please.PST.3SG
 ‘since the child liked the snowman after all’

¹ See Haspelmath (2013) for the term “person indexing” and its relation to traditional terms like “agreement” and “bound pronoun”.

Thus, there are many syntactic properties of clauses that do not depend on the verb, but those that do, viz. the verb's valency properties, are generally regarded as core features of syntax. Comparing valency properties across languages is thus a core concern of comparative syntactic research.

We can represent the valency of the English verb *accuse* and the German verb *gefallen* as in (8a–b). These representations are called **coding frames** in this work.

- (8) two exemplary coding frames
- a. English *accuse*: <Arg1-NOM V.SUBJ[1] Arg2-ACC of+Arg3>
 - b. German *gefallen*: <Arg1-DAT V.SUBJ[2] Arg2-NOM>

Here, Arg1, Arg2 and Arg3 are variables for distinct arguments that are primarily expressed by nominals; -NOM, -ACC and -DAT indicate the case features of the arguments, *of+* stands for a preposition that flags an argument, and V.SUBJ[] refers to the person information on the verb that matches the person feature of one of the arguments, i.e. the argument is indexed on the verb (in English, *accuse* indexes, or “agrees with”, its Nominative argument, and German *gefallen* also indexes its Nominative argument).

For a complete description of a verb's behaviour, its coding frame needs to be linked to a (possibly partial) representation of the verb's meaning in such a way that the argument variables are paired with variables for its semantic participants. We call such a semantic representation a role frame. The **role frames** of *accuse* and *gefallen* are shown in (9)–(10) with the linking to the valency frames:

- (9) English *accuse*
- a. role frame ‘X accuses Y of Z.’
(the accuser) (the accusee) (the wrongdoing)
 - b. coding frame <Arg1-NOM V.SUBJ[1] Arg2-ACC of+Arg3>

- (10) German *gefallen*
- a. role frame ‘X likes Y.’
(the liker) (the likee)
 - b. coding frame <Arg1-DAT V.SUBJ[2] Arg2-NOM>

To a significant extent, a verb's coding frame can be predicted from its role frame (cf. Levin & Rappaport Hovav 2005). For example, in English a participant with an agent role (e.g. accuser, breaker, thrower) is always linked to the Nominative argument in the coding frame. But not infrequently across languages, there are quite a few unpredictable linkings which simply have to be learned by speakers and listed in dictionaries. It is these unpredictable, or less predictable, coding frames that are

of greatest interest for the comparative study of valency, because they also tend to show the greatest cross-linguistic variability.

In addition to coding properties, verbs may also determine other properties of their arguments, such as the ability to be the antecedent of a reflexive pronoun, to be the target of omission under coreference, or the ability to be passivized. These properties are generally called **behavioural properties**. Arguments that are coded in the same way may nevertheless show different behaviours. For example, German *wiegen* ‘have a weight of (a measurement)’ takes an Accusative argument that cannot be passivized (cf. 11a–b), while *wiegen* ‘determine the weight of (an object)’ takes an Accusative argument that can be passivized (cf. 12a–b).

- (11) a. *Der Sand wiegt einen Zentner.*
 the sand weighs one hundredweight
 ‘The sand has the weight of one hundredweight.’
 (passive)
- b. **Ein Zentner wurde von dem Sand gewogen.*
 one hundredweight was by the sand weighed
- (12) a. *Frau Müller wiegt den Sand.*
 Ms. Müller weighs the sand
 ‘Ms. Müller determines the weight of the sand.’
 (passive)
- b. *Der Sand wird von Frau Müller gewogen.*
 the sand is by Ms. Müller weighed
 ‘The weight of the sand is determined by Ms. Müller.’

Thus, one can say that the two homonymous verbs *wiegen* have two different **valency frames**, perhaps notated as <Arg1/Subject V.SUBJ[1] Arg2/Extent> and <Arg1/Subject V.SUBJ[1] Arg2/Object>, respectively. In other words, we might use **syntactic-function** labels like **subject**, **object** and **extent** in valency frames, thus taking into account not only coding properties of arguments, but also behavioural properties. However, in our comparative study of valency, we have not been able to take these other properties into account systematically, so we generally use the label **coding frame** rather than **valency frame**. (Word order is intermediate between coding properties and behavioural properties; see § 11 below on word order.)

3 Notes on terminology

Before we go on to compare languages, we offer a few notes comparing linguistic traditions, especially with respect to terminology. The term “valency” was intro-

duced by Tesnière (1959) and has been used in English at least since Svartvik (1966), but other terms have been widely used as well.² The terms in (13) all have basically the same meaning. A terminological distinction between the more abstract “valency” and the more concrete “valency frame” is sometimes made, but is often redundant.

- | | | |
|------|-----------------------|---|
| (13) | a. complementation | e.g. Quirk et al. (1985: 1069–71) |
| | b. subcategorization | Chomsky (1965) |
| | c. argument structure | e.g. Goldberg (1995) |
| | d. government model | Mel’čuk (1974) (Russian <i>model’ upravlenija</i>) |
| | e. clause blueprint | Grebe (1959) (German <i>Satzbauplan</i>) |

Those nominals in a clause that are determined by (i.e. depend on the choice of) the verb are now generally called “arguments”, while the verb-independent elements are called “adjuncts”. Again, this distinction is widely recognized, but in the past the terminology varied quite a bit:

- | | | | |
|------|---------------|-------------------|-------------------|
| (14) | a. argument | adjunct | (here) |
| | b. complement | modifier, adjunct | e.g. Vater (1978) |
| | c. actant | circumstant | Tesnière (1959) |
| | d. argument | satellite | Dik (1997: 86–90) |

What we call the “role frame”, i.e. the semantic representation of the verb meaning with the participant variables, has been variously called “logical structure” (e.g. Van Valin & LaPolla 1997), “lexical-conceptual structure” (e.g. Jackendoff 1990), or “semantic valency”.

Throughout this chapter, we will work with the distinction between descriptive categories of particular languages and the comparative concepts of cross-linguistic research (Haspelmath 2010). In the practice of linguists, verbal valency is primarily used as a descriptive concept, needed to characterize the behaviour of particular verbs in particular languages. Thus, we need to develop a number of comparative concepts that allow us to compare valency patterns across languages.

The most important concepts are the semantic concepts **comparison meaning** (meanings that are used to compare verbs across languages, e.g. ‘miss’ and ‘fill’ in (1)–(2)) and **microrole** (meanings that are used to compare arguments with similar roles across languages, e.g. ‘liker’ and ‘likee’ in (10)), and concepts for formal coding elements, in particular **flags** (cases and adpositions) and **indexes** (see § 6 below). We have also tried to work with a comparative concept of argument, but as we will see in the next section, this is more difficult to apply consistently across languages than the comparison meanings, the microroles, and the coding elements.

² Note also that the alternative form “valence” is also common, especially in American English (e.g. Abraham 1978; Langacker 1988).

4 Delimiting valency: arguments and adjuncts

A lot of research on valency has been concerned with the practical task of compiling valency dictionaries, i.e. dictionaries of verbs that give a valency frame for every verb. This forces researchers to delimit the notion of valency very clearly, and to distinguish sharply between arguments (included in the valency frame of the verb) and adjuncts. But sometimes it is not quite clear whether a nominal is specific to the verb or not. Consider the bracketed phrases in (15).

- (15) a. *She put the book [on the table].*
 b. *He lives [in Wisconsin].*
 c. *She is sitting [on the sofa].*
 d. *They cut the meat [with the knife].*
 e. *He broke the window [with his fist].*

These are locational and instrumental phrases which could be taken either as arguments or as setting or manner adverbials and thus as adjuncts. Is there a way to tell whether they are arguments or adjuncts?

One way in which the distinction between arguments and adjuncts has often been framed is by considering the verb's meaning. If a participant is **entailed** by the meaning of a verb, this participant is an argument, otherwise it is an adjunct (e.g. Van Valin 2005; Bickel 2011). According to this criterion, *with the knife* in (15d) would be an argument, because *cut* means 'sever with a sharp instrument', while *with his fist* in (15e) would not be an argument, because *break* only means 'do something so that something becomes broken'. However, not all entailed participants can be overtly expressed, not all entailed participants would be considered arguments, and not all elements that are generally considered arguments are entailed participants.

First, not all entailed participants can be overtly expressed. For example, the English verb *tell* allows the expression of both the addressee and the content, but *lie* does not allow the expression of the content:

- (16) a. *She told her address to her friend.*
 b. *She lied (*her address) to her friend.*

Second, not all entailed participants would normally be considered arguments. For example, the verb *break* means 'do something₁ so that something₂ becomes broken'. The first entailed participant, the action that causes the breaking event, can be expressed in a *by*-phrase (*He broke the window by hitting it with his fist*), but such a *by*-phrase would not normally be regarded as an argument. More generally, all verbs that denote a spatiotemporal event entail a time and a location, but these are the most typical adjuncts, not arguments.

Third, some arguments are not entailed by the verb's meaning, e.g. the beneficiary *me* in (17).

(17) *Roland baked me a cake.*

As has been widely discussed, *bake* does not entail a beneficiary (one can bake a cake without having someone specific in mind), but the beneficiary is generally regarded as an argument.

Thus, because there is no close match between entailed participants and what are normally considered arguments, we do not adopt the entailment-based definition of *argument*.

We think that the notion that best captures the intuition that lies behind the argument-adjunct distinction is the notion of **verb-specificity**. Elements that are verb-specific are arguments, and elements that are not verb-specific are adjuncts. Perhaps the clearest case of verb-specificity is **coding-specificity**, i.e. where a verb determines idiosyncratic coding of its arguments, so that the coding is not predictable on the basis of the verb's meaning. This can best be seen in minimal pairs of verbs with very similar or identical meaning but different argument coding. Thus, the semantic role of *the boy* or *der Junge* is not very different in the (a) and (b) examples in (18)–(19), but only the (b) examples require a dative preposition or case. This is something that needs to be learned in addition to the meaning of the verb, and it is thus no accident that systematic valency research began in the context of language teaching (Helbig & Schenkel 1969).³

(18) English

- a. *I showed **the boy** the solution.*
- b. *I demonstrated the solution **to the boy**.*

(19) German

- a. *Ich unterstützte **de-n** **Junge-n**.*
I.NOM support.PST.1SG the-ACC boy-ACC
'I supported the boy.'
- b. *Ich half **de-m** **Junge-n**.*
I.NOM helped the-DAT boy-DAT
'I helped the boy.'

Another example of this type was given in the very first paragraph of this paper (English *wait for someone* vs. *expect someone*).

³ There was of course valency research before Helbig & Schenkel (1969), but they published the first valency dictionary, which was perhaps also the first book specifically on valency patterns in a language (incidentally, this work was carried out at Leipzig University and published in Leipzig).

Another fairly clear case of verb-specificity is **obligatoriness**. In many languages, some kinds of nominals must occur when a given verb is chosen. This is true not only of subjects (which are fairly trivial cases of arguments and therefore sometimes even ignored by valency researchers), but also of many objects, and sometimes even of oblique arguments. For example, the object in (20a) cannot be omitted (contrasting with 20b), and the prepositional *auf*-argument in (21a) cannot be omitted (contrasting with 21b).

(20) English

- a. *The dragon devoured **the princess**.*
(**The dragon devoured.*)
- b. *The customer ate (**the fish**).*

(21) German

- a. *Sie hat mich **auf den Fehler** hingewiesen.*
she has me on the mistake pointed
'She pointed me to the mistake.'
- a'. (**Sie hat mich hingewiesen.*)
- b. *Sie hat (**auf Godot**) gewartet.*
she has on Godot waited
'She waited for Godot.'

But there are many arguments that exhibit neither coding specificity nor obligatoriness. In fact, in many languages no argument is ever obligatory, because all arguments can be omitted when they can be reconstructed from the context.⁴ And in many cases, the absence of an argument may be odd for pragmatic reasons, not for any grammatical reasons:

- (22) a. ??*He lives.*
- b. ??*She is sitting.*

Hearing about someone living or sitting is not informative, so the location is usually specified to get a pragmatically felicitous utterance. Thus, obligatoriness is not as useful for delimiting valency as is often thought, not even for those languages where it does play a role.

⁴ However, often one can make a clear distinction between an anaphoric and an existential interpretation of argument omission, and when argument absence implies an anaphoric interpretation, this could be taken as evidence of verb-specificity and argumenthood. Thus, in *She found out* the omitted argument has to be definite and anaphoric (= *She found out about it*), whereas in *She ate* the omitted argument has to be indefinite and non-anaphoric (= *She ate something*).

The crucial criterion for argumenthood is whether a nominal is limited in its cooccurrence options to a restricted and semantically arbitrary set of verbs (i.e. whether it is verb-specific), or whether it can occur with any verb, or at least with a large and semantically coherent class of verbs (i.e. whether it is verb-free). We can call this **specificity of occurrence**. Thus, while the boldfaced arguments in (23) are not coded in a (highly) verb-specific way and are not obligatory, they are still considered arguments because they cannot occur freely with any verb, as shown in (24).

- (23) a. *He called **his brother**.*
 b. *She gave **the shopkeeper** too much money.*
 c. *He ran **to the house**.*
- (24) a. **She laughed **her brother**.*
 b. **He spent **the shopkeeper** too much money.*
 c. **She sang **to the house**.*

One widely cited test for (non-)argumenthood that reflects specificity of occurrence is the ‘happen’ test: If a sentence has a paraphrase in which a phrase from the original sentence is removed and occurs in an anaphoric ‘happen’ (or ‘do so’) clause, then this phrase is an adjunct:

- (25) a. *She called her brother **in the morning**.*
 b. *She called her brother, and this happened **in the morning**.*
- (26) a. *They accused her of plagiarism **in her school**.*
 b. *They accused her of plagiarism, and this happened **in her school**.*
 (OR: *and they did so in her school*)

This test shows clearly that temporal and locational setting adverbials are adjuncts, while obligatory and coding-specific nominals are arguments:⁵

- (27) a. *She called **her brother**.*
 b. **She called, and this happened **her brother**.*
- (28) a. *They were waiting **for Godot**.*
 b. **They were waiting, and this happened **for Godot**.*⁶

⁵ In the ValPaL manual, we asked the contributors to determine argumenthood by means of a test of this kind. But several participants told us that no such test is available in their language (e.g. Seifart 2013 on Bora). Thus, this test is not readily available as a cross-linguistically applicable way of determining arguments (see also Haspelmath 2014).

⁶ This sentence is acceptable with an irrelevant sense (‘this happened for the sake of Godot’).

The test can be applied to many languages, because anaphoric verbal expressions like ‘do so’ or ‘this happened’ are widely found. However, with (stationary)⁷ locational phrases and instrument nominals, the result of the test is not so clear:

- (29) a. *She was sitting **on the sofa**.*
 b. *?She was sitting, and this happened **on the sofa**.*
- (30) a. *They cut the meat **with the knife**.*
 b. *?They cut the meat, and this happened **with the knife**.*

Quite generally, locational phrases and instrument nominals are hard to classify uniquely as arguments or adjuncts. The concept of valency and the argument/adjunct distinction is simply not very useful for these kinds of phrases.⁸

In addition to coding specificity and obligatoriness, quite a few other language-particular criteria for distinguishing between arguments and adjuncts have been cited for different languages in the literature (cf. Haspelmath 2014). However, since we are interested in cross-linguistic comparison, language-particular argument criteria are not of interest to us.

Thus, for quite a few cases we did not have a unique way of distinguishing between arguments and adjuncts, and the ValPaL database is therefore not consistent in this regard. If the verb ‘break’ is said to have three arguments in one language and two arguments in another language, this does not mean that there is a typological difference between the two languages. This may just be due to an arbitrary or at least idiosyncratic difference between the decisions taken by the authors. As a result, the number of arguments is not a kind of information that should be taken as important for cross-linguistic comparison, especially when the difference concerns locational phrases and instrument nominals.

5 Comparison meanings: verbs and microroles

For cross-linguistic comparison of valencies, or more specifically coding frames, we need to be able to compare verbs (the valency-bearers), participant roles, and coding elements (flags and indexes, see § 6 below). For example, a comparison as in (2a–b), repeated here from above, presupposes that we know that both English

⁷ By contrast, directional locational phrases are clearly arguments: *He ran to the house* / **He ran, and this happened to the house*.

⁸ Note also that the ‘happen’ test cannot be applied to subjects in English; and the criterion of verb-specificity might be taken to indicate that English subjects are not arguments, because they occur with all verbs. (However, nobody has suggested that subjects are not arguments, in English or other languages, so this has no practical consequences.)

fill and Chintang *phatt-* mean ‘fill’, that the verbs in addition to an agent role have a substance role (in this example, the water) and a location role (in this example, the container), and that the substance role is coded by a preposition (*with*) in English, and with Nominative case in Chintang.

- (31) a. English
She₁ filled the container₂ with water₃.
- b. Chintang (Tibeto-Burman, Nepal)
Hũsa-ŋa₁ cuwa-Ø₃ gagri-be₂ phatt-e.
 she-ERG water-NOM container-LOC fill-PST
 ‘She filled the container with water.’
 (Literally ‘She filled the water into the container.’)
 (Bickel et al. 2010: 387)

Thus, our ValPaL database contains entries such as those in (32a–b). The first line contains the verb form (its citation form), the second line contains the coding frame, and the third line contains the microroles, i.e. the participant roles relevant for this particular verb.

- (32) a. English *fill*
 <1-NOM V.subj[1] 2-ACC with+3>
 (1=filler, 2=filled container, 3=filling material)
- b. Chintang *phatt-*
 <1-ERG V.subj[1] 2-NOM 3-LOC>
 (1=filler, 3= filled container, 2= filling material)

In other words, the verb entries in our database contain the kinds of information that we saw above in (9)–(10), though in a slightly abbreviated notation, with microrole labels instead of full role frames with index numbers (thus, “1=filler, 2= location, 3=substance” is equivalent to “Arg1 fills Arg2 with Arg3”).

The 80 pre-defined comparison meanings that we used to gather comparable verb forms for the database are identified by their English counterpart (which is also their label) and by a typical context. For example, the meaning ‘cover’ is identified by the label COVER as well as the typical context “The woman covered the boy with a blanket.”. We did not try to provide a more detailed semantic description, as we did not expect this to lead to greater comparability of verbs across languages.⁹

⁹ We are aware that the use of English labels to elicit counterpart verbs, as well the use of English as a general metalanguage of our project, may well have biased the set of comparison meanings toward the kinds of meanings that tend to have simple expression in English and related languages. Unfortunately, we felt that there was nothing we could do about this, other than avoiding verb

Languages differ not only in their forms, but also in the kinds of meanings that they have words for, so the verbs that are given as counterparts to our 80 comparison meanings are not always perfect matches. For this reason, we do not call them “equivalents”, but “counterparts”. We asked the contributors to find the semantically closest verb in their language. Moreover, we said that the verb should have a “basic” flavour, i.e. verbs that are used very rarely should be avoided if a more common verb with similar meaning is available. Sometimes a basic verb may have a somewhat different meaning from the English label that serves as comparison meaning; in such cases, we said that deviation from the comparison meaning was less important than basicness. The reason for this was that we felt that there was no need to insist on exact matches, because this was not achievable anyway in many situations. And it must be kept in mind that the purpose of our database is not to provide exact translations from each language to each other language, but to compare languages with respect to their valency patterns.¹⁰

The relationship between the pre-defined comparison meanings and counterpart verbs can be many to many, so that if there were two basic verbs in the language that corresponded to a given meaning, both could be included. For example, in Sliammon, there are two verbs corresponding to EAT, *ʔittən* and *məkʷt*, with two different coding frames (Watanabe 2013). Conversely, when a single verb corresponded to two different meanings, it was not necessary to enter it twice into the database. For example, the Mandinka verb *bori* is a counterpart both of RUN and of ROLL (Creissels 2013). Homonymous verbs are distinguished by a number, e.g. Nllng *nllaa* (1) (‘dress’) and *nllaa* (2) (‘live’).¹¹

The argument variables in the coding frames are represented by integers (1, 2, 3, ...) whose order does not have any significance (though normally the linearly first argument gets variable 1, and so on). Each argument variable is linked to a microrole, so that we know what role the argument plays in the verb meaning. With typical transitive verbs, as in (33), the microroles are not so important, because they could normally be replaced by “agent” and “patient”, and with single-argument verbs, as in (34), they are completely redundant.

meanings for which we were aware from the beginning that there is something peculiar about their English counterparts.

10 Every translation distorts the original text to some slight extent, but translations are still eminently useful. We do not expect our cross-linguistic comparison to be any better (or worse) than the average professional translation.

11 Since a verb must have a unique coding frame in our database (this was a decision we took to simplify the database), if a verb had two slightly different meanings but different coding frames, it had to be broken up into two different verb entries that are treated as homonymous. For example, Yaqui *chaae* has the coding frame <1-NOM V> when it renders ‘scream’, but <1-NOM 2-ACC-DIR V> when it means ‘shout at’ (Estrada-Fernández et al. 2013). ‘Scream’ and ‘shout’ would be sufficiently similar to count as one meaning, and no doubt most linguists would say that we are dealing with the same verb here, but since there are two possible coding frames, two different ValPaL database entries are required.

- (33) a. EAT eater eaten food
 b. WASH washer washed entity
 c. HELP helper helpee
- (34) a. COUGH cougher
 b. SINK sunken entity

(Microroles sometimes have unusual-sounding labels such as *helpee*; we hope that readers and database users will quickly get used to them. We find transparent labels much more practical than opaque abbreviations or numbers.)

The microroles are important for those verb meanings where cross-linguistic metataxis is not uncommon. We saw some examples in (1)–(2/31) above, and here is another contrasting pair of verbs that are counterparts of the same meaning from two languages:

- (35) Sri Lanka Malay (Nordhoff 2013)
 a. verb: *mintha-* ‘ask for’
 b. coding frame: <1 2-ACC 3-LOC V>
 (1=asker, 2=requested thing, 3=askee)
 c. example:
 see baapa=ka car=yang su-mintha
 1SG father=LOC car=ACC PST-beg
 ‘I asked my father for the car.’ (Lit. ‘I asked the car at my father.’)
- (36) Jaminjung (Schultze-Berndt 2013)
 a. verb: *yanggi ganarrany* ‘ask for’
 b. coding frame: <1-ERG 2-ABS 3-DAT SBJ[1].OBJ[2].V>
 (1=asker, 2=askee, 3=requested thing)
 c. example:
 gurrany yanggi ya-wun-karra=yinyag mangarra-wu
 NEG ask IRR-DU.A:1.P-put=1DU.EXCL.P plant.food-DAT
 ‘Don’t ask us two for food, you two!’ (Lit. ‘Don’t ask us to food.’)

The microroles ‘askee’ and ‘requested thing’ do not clearly map onto the general roles ‘patient’, ‘source’ and/or ‘recipient’, but in our approach this does not matter: The coding frames in (35b) and (36b) are perfectly clear and comparable on the basis of the microroles. Intuitively, the Jaminjung sentence is more like English (using a dative case or the requested thing), while the Sri Lanka Malay sentence is literally ‘I asked the car at my father’, which is rather different from English, and more like Russian *poprosit’* (*ja poprosila mašinu u otca* [I asked the.car at father]). So there is a metataxis relationship between (35) and (36).

6 Coding elements

We saw above (§ 2) that valency has both a coding and a behavioural component, but here we limit ourselves to coding frames. And while behavioural properties of arguments are quite heterogeneous, there are just two kinds of **coding elements**, which are quite straightforward to represent schematically: **flags** (adpositions or case markers) on arguments, and **indexes** on the verb that match arguments. Let us consider three additional examples, which illustrate the conventions that we use:

(37) Japanese (standard) (Kishimoto & Kageyama 2013)

- a. verb: *kowasu* ‘break’
- b. coding frame: <1-NOM (2-INSTR) 3-ACC V>
- c. example:
Ken-ga (hanmaa-de) mado-o kowasi-ta.
 Ken-NOM hammer-INS window-ACC break-PAST
 ‘Ken broke the window (with a hammer).’

(37) Hoocak (Hartmann 2013)

- a. verb: *hožu* ‘load’
- b. coding frame: <1 2 LOC3 UND[2].ACT[1].V>
- c. example:
wiiraruti=ra kšee ho<ha>žu
 trailer=DEF apple <1E.A>put.in
 ‘I loaded apples on the trailer.’

(39) Icelandic (Barðdal 2013)

- a. verb: *stela* ‘steal’
- b. coding frame: <1-NOM V.AGR[1] 2-DAT (*frá*+3-DAT)>
- c. example:
Þjófur-inn stal peningum frá gömlu konu-nni.
 thief-the.NOM stole money.DAT from old.DAT woman-the.DAT
 ‘The thief stole money from the old lady.’

The following conventions are used to represent coding elements (flags and indexes) in coding frames in ValPaL:

- (i) Cases are represented by (abbreviations of) their category labels, such as NOM (for nominative), DAT (for dative), following the argument variable and linked to it by a hyphen.

- (ii) Adpositions are represented by their form, linked to the argument variable by a plus (+) sign, preceding it (for prepositions) or following it (for postpositions).
- (iii) Index-sets are represented by their category labels, linked to the verb by a period (e.g. V.AGR in Icelandic). The index label is immediately followed by brackets which contain the argument variable number, so “V.AGR[1]” in Icelandic means that the agreement index-set corresponds to the argument that also bears nominative case (the stealer in the case of the verb *stela* ‘steal’), and “UND[2].ACT[1].V” in Hoocak means that arguments 1 and 2 are indexed by the Actor and Undergoer sets, respectively (the loader and the loaded theme in the case of *hožu* ‘load’).
- (iv) Optional arguments may be enclosed in brackets. (We did not enforce this, because optionality is a difficult concept, and we were not able to provide this information consistently.)

7 Kinds of arguments

In the most common case, verbal arguments are nominals (noun phrases, adpositional phrases) or person indexes on verbs. Such arguments are represented in our coding frames by free-standing integers (possibly accompanied by adpositions or with case labels attached to them) or by integers in brackets following the index-set label which is attached to the verb variable.

But there are certain other kinds of arguments, in particular locational arguments, clausal arguments, and utterance arguments.

A locational argument need not have any particular categorial form. It is most often expressed by an adpositional phrase, but the nature of the adposition is determined by the spatial meaning, not by the verbal valency. In many languages, locational adverbs may also be expressed by spatial adverbs which are neither adpositional phrases nor noun phrases. For example, a locational argument occurs with Italian *sedere* ‘sit’, illustrated in (40a–c).

(40) Italian (Cennamo & Fabrizio 2013)

- a. *Mario siede in seconda fila.*
Mario sits in second row
‘Mario is sitting in the second row.’
- b. *Gli anzian-i sede-va-no intorno a-l fuoco.*
the old.man-PL sit-IPFV-3PL around to-the fire
‘The old men were sitting around the fire.’
- c. *Mario sied-e lì.*
Mario sits there
‘Mario is sitting there.’

In ValPaL's coding frames, locational arguments are notated with LOC, followed by the argument number. Thus, the coding frame of Italian *sedere* is <1 V.SUBJ[1] (LOC2)>. (A locational argument also occurs in the coding frame of Hoocak *hožu* 'load', see (38b) above.)

Clausal arguments are complement-clause arguments such as *that*-clauses and infinitival clauses in English (e.g. *I hope that you'll be there; I hope to be there*). These were excluded from ValPaL, as the typology of complement clauses is an entirely separate domain of study. The only comparison meanings that frequently require a clausal argument are 'know' and 'want', as illustrated by (41b). However, we asked our contributors for contexts like (41a), where the 'wanted thing' argument is a nominal rather than a clause. Thus, Eastern Armenian *uzel* 'want' has the coding frame <1-NOM 2-NOMDAT V.SUBJ[1]> in ValPaL.

(41) Eastern Armenian (Khurshudian & Daniel 2013)

a. *Jes uz-um em ajs aḡžka-n.*
I.NOM want-CVB.IPFV AUX.1SG this girl.DAT-DEF
'I want this girl.'

b. *Jes uz-um em gn-al.*
I.NOM want-CVB.IPFV AUX.1SG go-INF
'I want to go.'

There is another comparison meaning that takes a special kind of argument: 'say' usually takes what we call an **utterance argument** (and marginally 'tell' and 'ask' can have such an argument as well). Like locational and clausal arguments, utterance arguments cannot be treated like nominal arguments because they are not coded by the usual flags (cases, adpositions) or indexes. Most commonly, they are simply juxtaposed (as in 42a–b), but occasionally there are special quotative markers, as in Japanese (see 42c).

(42) a. Jakarta Indonesian (Connors & Gil 2013)

Nadia omong kita mendingan pergi.
Nadia say 1PL better go
'Nadia said it would be best if we go.'
(OR: 'Nadia said: "We better go".')

b. Ket (Vajda & Kryukova 2013)

Qarⁱ da'ŋa bara: "eddi".
DEM.M 3SG.M.DAT 3.M.SG.say alive-1SG
'He said to him: "I'm alive".'

c. Japanese (standard) (Kishimoto & Kageyama 2013)

Ken-ga Mari-ni "hai" to it-ta.
Ken-NOM Mari-DAT yes QUOT say-PST
'Ken said "yes" to Mari.'

Utterance arguments are notated with UTT in ValPaL, followed by the argument number (thus, Jakarta Indonesian *omong* has the coding frame <1 V (UTT2)>).

8 Prominence-based splits

Coding frames contain a unique coding element, i.e. a unique case or adposition associated with a free-standing argument number, or a unique index-set on the verb that indexes an argument. But sometimes languages exhibit prominence-based splits, i.e. depending on the argument's inherent or contextual prominence features, different flags are employed. A well-known case of this is differential object marking in Spanish, where human direct objects have the preposition *a* (e.g. *veo a Juan* 'I see Juan'), while nonhuman objects lack this preposition (e.g. *veo la casa* 'I see the house'). In such cases, one would not say that the verb *ver* 'see' is associated with two different coding patterns, because the coding contrast does not depend on the verb, but on the nature of the argument.

But an even better-known case is English, where a Nominative-Accusative distinction is made only with personal pronouns (*I/me, he/him, she/her, we/us, they/them, who/whom*), i.e. with nominals that are higher in prominence. All other nominals make no distinction. When a distinction is only partial and depends on the kind of nominal, then the usual solution adopted by linguists is to encode the distinction in the syntactic rule and to make a separate (morphological) statement about the lack of the distinction in certain kinds of nominals. Thus, in Russian we would say that there is a general Nominative-Accusative distinction, even though it is syncretized in many inanimate nouns in the singular (e.g. NOM/ACC *kniga/knigu* 'book', but *stol/stol* 'table', *mesto/mesto* 'place'), and in all inanimate nouns in the plural (e.g. NOM/ACC *knigi/knigi* 'books'). Likewise, we decided to say that English has a general Nominative-Accusative distinction, even though it is syncretized in all nominals that are not personal pronouns (and in the personal pronouns *you* and *it*). Thus, the English verb *see* has the coding frame <1-NOM > V.SUBJ[1] > 2-ACC> in ValPaL.

Another case of a prominence-based split is Eastern Armenian, where a direct object is consistently in the Dative case when it is human, but in the Nominative case when it is inanimate (Khurshudian & Daniel 2015). In the glosses in the Eastern Armenian examples, the case labels DAT and NOM are used, e.g.

(43) Eastern Armenian (Khurshudian & Daniel 2013)

- a. *Mard-ə tesa-v tva-ji-n.*
 man[NOM]-DEF see.AOR-AOR.3SG boy-DAT-DEF
 'The man saw the boy.'
- b. *Jes c^hamak^h em tesn-um.*
 I.NOM land[NOM] AUX.1SG see-CVB
 'I see the land.'

However, since the Dative and the Nominative occur in complementary distribution with direct objects, they are treated as a single case (“Dative-Nominative”) for the purposes of valency information, and the coding frame of *tesnel* ‘see’ in the database is <1-NOM 2-NOMDAT V.SUBJ[1]>.

9 Verb forms

Following our guidelines, the verbs in our database have to be conventional expressions in the language, but they do not have to be basic or simple verbs, and they do not even have to be “verbs” in a very strict sense.

The requirement that the entries are conventional (or FIXED) expressions means that we did not want the contributors to provide ad-hoc paraphrases. For example, in Bezhta the verbal meaning ‘peel’ is rendered by ‘take (off)’ plus ‘skin’ (Comrie & Khalilova 2013).

- (44) *Kibba k’atu-wa-s beš b-ayo-yo.*
 girl.ERG potato-OBL-GEN1 skin(III).ABS III-take-PST
 ‘The girl peeled the potato.’ (Lit. ‘The girl took (off) the skin of the potato.’)

Another example is the meaning ‘sing’, which is rendered by *donkilóo láa* [song say] in Mandinka. This is a fully regular combination of a verb (‘say’) and its object (‘song’). Other objects would be equally possible (e.g. ‘tell a story’). Thus, the database has “no counterpart” for SING in Mandinka and PEEL in Bezhta.

Since it is very difficult to distinguish between fixed expressions and freely created expressions, our database may not be homogeneous in this regard. For example, for ‘rain’, we have *bi-ga o-da* [rain-NOM come-DECL] in Korean, but in Bezhta, the very similar *wodo guu-s* [rain come-PRS] was not counted as a conventional expression. This may not correspond to a real difference in the fixedness of the two expressions.

The counterparts of our comparison meanings are not always really verbs in the languages, but we wanted to avoid the issue of the verb-adjective distinction and thus basically ignored the language-particular status of the counterparts. We were exclusively interested in the valency properties of expressions corresponding to meanings such as ‘be afraid’, ‘like’, ‘feel cold’, ‘be sad’, ‘be dry’, not in their word-class assignment. It is apparently the case that all counterparts of dynamic comparison meanings are actually verbs in all languages, but counterparts of meanings such as ‘be afraid’ and ‘be sad’ (and especially ‘be dry’) are very often called “adjectives” in language descriptions. In such cases, the entries often include copulas, but these copulas should be regarded as irrelevant to our underlying

ing question of valency (we did not try to be consistent in including or excluding copulas).¹²

While fully regular ad-hoc paraphrases are not included, as just noted, **complex verbal expressions** are allowed in the database, and were in fact used in many cases (more than a quarter). They may be quite diverse formally: They may consist of a verb plus an additional locative particle (e.g. English *sit down*), or a verb plus a reflexive pronoun (e.g. German *sich hinsetzen* ‘sit down’), or a verb plus an incorporated noun (e.g. Bezhta *komak bowal* ‘help’, lit. ‘do help’), or two verbs (e.g. Yoruba *mú ... wá* ‘bring’, lit. ‘take ... come’). They may also include valency-changing affixes, especially causative affixes, e.g. Mandinka *niki-ndi* ‘teach’, which derives from *nikin* ‘learn’.

It is important to be aware that by including complex verbal expressions in the database, we are on the one hand maximizing the amount of data that we can include for each language, but on the other hand we are introducing a possible bias. The comparison meanings that we started out with almost all have simple counterparts in English (and closely related European languages), but not uncommonly, they do not have simple counterparts in other languages. Thus, the English verb set (Goddard 2013) consists of simple verbs, while the verb sets of other languages often include derived verbs such as causatives or other complex verbal expressions. Since complex verbal expressions may have different valency profiles than simple verbs, we may not get an unbiased picture of the general valency character of the language.

We chose to include complex verbal expressions in order to maximize the amount of data in the database, and because it is not clear that complex verbal expressions are less fundamental to a language’s character than simple, monomorphemic verbs. The question of how best to compare verbal syntax across languages in a systematic way is still quite open, in our view, so we did not want to limit the available data in advance. We did try to annotate all verbs in the database for simple vs. complex status,¹³ so that users can filter out complex verbal expressions if they want.

With complex verbal expressions that include a nominal element, the question may arise what the relation of this nominal element is to the verb’s valency. Consider the examples in (45)–(47).

¹² One of our comparison meanings was ‘be a hunter’. Including this was motivated by the question whether the subject of the predicate ‘is a hunter’ is coded in the same way as the subject of other stative predicates such as ‘be sad’ or ‘be afraid’. We recognize that including such a meaning in a database on verbs is somewhat confusing, and as a result some contributors did not include a counterpart.

¹³ More precisely, a *complex* verb in the database is a verbal expression that includes an element that is relevant for the verb’s valency, such as a causative marker, an applicative morpheme or an incorporated noun.

- (45) Bezhta *komak bowal* [help do] ‘help’
Kibba iyo-l komak b-oo-yo.
 girl.ERG mother-LAT help(III) III-do-PST
 ‘The girl helped her mother.’
 (Lit. ‘The girl did help to the mother.’)
 (Comrie & Khalilova 2013)
- (46) Yucatec Maya *ch’a’ sahakil* [take fear] ‘be afraid’
Le wíinik-o’ t-u ch’a’-ah sahak-il ti’ le báalam-o’.
 DEM man-D2 PFV-SBJ.3 take-CMPL fear LOC DEM jaguar-D2
 ‘The main became afraid of the jaguar.’
 (Lehmann 2013)
- (47) Icelandic *finna lykt* [find smell] ‘smell’
Björn-inn fann lykt-ina af strá-k-num.
 bear-the.NOM found smell-the.ACC of boy-the.DAT
 ‘The bear smelled the boy.’
 (Lit. ‘The bear found the smell from the boy.’)
 (Barðdal 2013)

In these cases, one might at first want to say that we are dealing with transitive verbs, and that the incorporated nouns (‘help’, ‘fear’, ‘smell’) are objects of these verbs. This would not be incorrect, and indeed at some level this needs to be said, because otherwise the case-marking (ergative marking of the helper in (45), accusative marking of the incorporated noun in (47)) and the agreement (gender III agreement of the verb with ‘help’ in (45)) could not be explained. Certainly the Bezhta ‘do’ verb, the Yucatec Maya ‘take’ verb and the Icelandic ‘find’ verb are transitive, and this explains some facets of the complex verbal expressions. But from the comparative perspective, it is these complex verbal expressions as a whole which are the counterparts of the comparison meanings, and thus their valencies are as in (48a–c).

- (48) a. Bezhta *komak bowal* [help do] ‘help’
 <1-ERG 2-LAT V>
 (1=helper, 2=helpee)
- b. Yucatec Maya *ch’a’ sahakil* ‘be afraid’
 <SBJ[1].V 1 le+2>
 (1=fearer, 2=fear stimulus)
- c. Icelandic *finna lykt* ‘smell’
 <1-NOM V.AGR[1] af+2-DAT>
 (1=smeller, 2=smelled entity)

These “composed valencies” may sometimes look a bit odd, especially for Bezhta: A verb with an ergative but no absolutive argument is not normally possible, and most verbs normally agree with one argument in gender. In purely language-specific descriptions, composed valencies are often left aside, but strictly speaking, they are necessary for language descriptions, too. That the action of helping is standardly expressed by ‘do help’ in Bezhta is not fully predictable (alternatively, one could say ‘give help’, ‘extend help’, etc., or there could be a completely unrelated simple verb), and that the smelled entity is coded by the preposition *af* in Icelandic is not fully predictable (cf. also English *take part*, where it is not predictable that the action is coded with the preposition *in*: *take part in something*).

Thus, the general principle of the ValPaL database is that only arguments that can be varied freely are taken into account as variables in the coding frames. Elements which are fixed parts of complex verbal expressions are ignored. In the coding frame <1-ERG 2-LAT V> of Bezhta *komak bowal*, the variable V stands for the complex verbal expression. The internal structure of counterpart verbs (or verbal expressions) is not reflected in the database.

10 Argument mismatches: apotaxis

In the simplest case, which is very frequent in the ValPaL database, the arguments of the counterpart verbs have roles that correspond to the microroles that we anticipated, and our anticipated microroles have corresponding arguments.

But this is not always the case. What we anticipated may admittedly have been based on English and similar European languages. But languages may differ not only in the way in which arguments expressing the same roles are coded (this is metataxis, cf. § 1 above). They may also differ in the roles that can be expressed as arguments (this can be called “apotaxis”). Let us illustrate this with the verb meaning ‘dig’, which involves the creation of a hole (i) in the ground (ii) by a digger (iii), possibly with the goal of digging up a thing (iv) that was hidden in the ground. (In addition, digging of course requires an instrument, minimally a body part, but since the coding of instruments exhibits little variability, this is left aside here.) In German, the digger and the ground can be expressed simultaneously with the hole or with the hidden thing:

(49) German

- a. *Sie gruben ein tiefes Loch in die Erde.*
they dug a deep hole in the.ACC earth
‘They dug a deep hole in the earth.’
- b. *Sie gruben in der Erde nach dem Schatz.*
they dug in the.DAT earth after the treasure
‘They dug for the treasure in the earth.’

But the hole and the hidden thing cannot be expressed simultaneously (**Sie gruben ein Loch nach dem Schatz* ‘They dug a hole for the treasure’).

In other languages in our database, our contributors gave us verbs that take primarily the ground as the argument of digging (e.g. Bezhta in 50, Hoocak in 51), or primarily the hidden thing (e.g. Bora in 52).

- (50) Bezhta (Comrie & Khalilova 2013)
Öždi mäche b-äx-čä.
 boy.ERG ground(III).ABS III-dig-PRS
 ‘The boy digs the ground.’

- (51) Hoocak (Hartmann 2013)
Maq=ra ha-k’e.
 earth=DEF 1E.A-dig
 ‘I’m digging the ground.’

- (52) Bora (Seifart 2013)
Mújcuri ó tsehdí.
 barbasco.root(ACC) 1SG dig
 ‘I dig for barbasco root.’

Another manipulation verb meaning that shows variability is ‘peel’. In Russian, both the peel and the peeled object can be expressed as arguments (as seen in 53), while in German, only the peeled object can be an argument (at least with the verb *schälen*, 54). In Yucatec Maya, finally, only the peel is an argument (as seen in 55).

- (53) Russian (Malchukov & Jahraus 2013)
Povarënok očisti-l kartošk-u ot kožur-y.
 kitchen.boy.NOM peel.PFV-PST.M.SG potatoes-ACC from skin-GEN
 ‘The kitchen boy peeled the potatoes.’
 (Lit. ‘The kitchen boy cleaned the potatoes of their skin.’)

- (54) German
Die Großmutter schälte die Kartoffeln.
 the grandmother.NOM peeled the potatoes.ACC
 ‘Grandmother peeled the potatoes.’

- (55) Yucatec Maya (Lehmann 2013)
Le xibpal-o’ t-u lak-ah u sóol le che’-o’.
 DEM boy-D2 PFV-SBJ.3 detach-CMPL POSS.3 shell DEM wood-D2
 ‘The boy removed the bark from the stick.’
 (Lit. ‘The boy removed the stick’s shell.’)

These differences between languages must be kept in mind when interpreting the cross-linguistic data of the ValPaL database. While most verbs have arguments that correspond largely to arguments in other languages, there are sometimes apotactic mismatches of the type just seen.

Another example is the comparison meaning ‘shave’. A verb with this meaning can take the shaved person as its non-agent argument (e.g. in Chatino, in 56), or the body part (e.g. in Ainu, in 57), or the hair/the beard (e.g. in Balinese, in 58), or it can be intransitive, because the affected entity is understood from the context (e.g. in Xârâcùù in 59, and also in English).

(56) Zenzontepec Chatino (Campbell 2013)

Yū=wá nka=téē? =yu j=yū.
 3SG.M=DEM CPL.CAUS-get.shaved=3SG.M OBL=3SG.M
 ‘He shaved (lit. himself).’

(57) Ainu (Bugaeva 2013)

E=nan-u hu memke.
 2SG.A=face-POSS shave
 ‘Shave (lit. your face).’

(58) Balinese (Shibatani & Artawa 2013)

Anak=e muani ento nguris jenggot=ne.
 person=DEF male that shave beard=3.POSS
 ‘The man shaved (lit. his beard).’

(59) Xârâcùù (Moyses-Faurie 2013)

Nâ xii.
 he shave
 ‘He shaved.’

Similar cases of apotaxis are also found with other verbs involving the human body, such as ‘hurt’, where either the affected person can be an argument of the verb (as in English *I feel pain*), or the affected body part (as in English *My head is hurting*).

In general, apotactic variability in valency patterns can be attributed to metonymic shift (e.g. Waltereit 1998), but there is a lot of interesting cross-linguistic variation here that deserves further study (see also Michaelis & APiCS Consortium 2013 on diverse ways of expressing having a headache in pidgin and creole languages).

11 Constituent order

Unlike argument flagging and argument indexing, constituent order is not a coding element that is generally available. The primary way in which constituent order may serve to code semantic roles in some languages is by requiring agents to precede the verb and patients to follow the verb, as in typical SVO languages like English or Yoruba. Since constituent order is salient in some languages, we allowed the contributors to provide this information, but since it is often difficult to say whether a particular order is required for coding a relation, we made this an optional feature. In (60) we give a few examples of coding frames that include constituent order information.

- (60) English
 a. *be afraid* ‘fear’ <1-NOM > V.SUBJ[1] > of+2>
 Emai
 b. *dia* ‘live’ <1 > V > vbi+2>
 Mandarin Chinese
 c. *zhuāng* ‘load’ <1 > (LOC2 >) V > 3>
 Mandinka
 d. *ma* ‘touch’ <1 > 2 > V > 3+la>

Mandarin Chinese and Mandinka are atypical in that they have fairly rigid constituent order, but have some arguments that must occur in preverbal position, while others must occur in postverbal position:

- (61) Mandarin Chinese (Zhang Guohua 2013)
Gōngrén-mén wǎng kǎchē-shàng zhuāng gàncǎo.
 worker-PL PREP truck-LOC load hay
 ‘The workers loaded hay onto the truck.’
- (62) Mandinka (Creissels 2013)
Kambaan-óo ye saá maa fál-oo la.
 boy-DEF PFV.POS snake.DEF touch stick-DEF OBL
 ‘The boy touched the snake with a stick.’

While constituent order is represented in the coding frames of these languages in ValPaL, order is not regarded as a coding element.

12 Alternations

In many languages, verbs may systematically be associated with different valencies under different circumstances. Such situations are called **valency alternations**.

Three well-known examples from English are the Passive alternation, the Dative alternation and the Causal-noncausal alternation (also known as “ambitransitive” or “labile” verbs).

(63) Passive ($V \rightarrow be V'-en$)

<1-NOM V 2-ACC> \rightarrow <2-NOM V' (by+1)>

- a. *The bear caught a fish.*
- b. *A fish was caught by the bear.*

(64) Dative alternation (uncoded)

<1-NOM V 2-ACC to+3> \rightarrow <1-NOM V' 3-ACC 2-ACC>

- a. *She gave the money to her brother.*
- b. *She gave her brother the money.*

(65) Causal-noncausal alternation (uncoded)¹⁴

<1-NOM V 2-ACC> \rightarrow <2-NOM V'>

- a. *He opened the door.*
- b. *The door opened.*

Like valency frames, valency alternations generally do not affect all verbs equally and thus subclassify the verbal lexicon in a language. Alternations often introduce many additional complexities and thus could not be covered exhaustively in ValPaL. The general guideline was that not more than ten alternations had to be entered if the language had more. Also, alternations which more or less apply across the board (as in some languages causatives do) did not have to be included either, as they would not be very useful in distinguishing verb classes.

Alternations are sometimes subdivided into voices (like passive, middle and antipassive) and valency-changing operations (like causative, applicative, desubjective), but this distinction cannot be made consistently across languages. It is sometimes correlated with inflectional vs. derivational status, but this distinction likewise cannot be made consistently across languages. Another frequently made association is between voice and information-structural function, but other alternations may also have informational-structural effects (e.g. Peterson 2007: Chapter 4). Thus, we do not distinguish between voices and other valency-changing alternations.

Alternations can be grouped into **coded alternations**, where the verb undergoes a change (as in the English Passive alternation), and **uncoded alternations**, where the form of the verb is the same with the two alternate valency frames (as in 65 for English). In addition, alternations may preserve the semantic roles, or

¹⁴ This alternation has often been called “inchoative-causative” alternation. See Haspelmath et al. (2014) for the more general and more transparent term “causal-noncausal”.

they may change the available semantic roles (by removing a role as in (65), or by adding a role as in (66) below), as long as the same basic meaning of the verb is preserved.

When an alternation was unproductive and occurred only with a small number of verbs, we were less interested in the alternation. (When the valency-frame variation affects only a single verb, it is not really an alternation, because alternations require some systematicity.)

In (63)–(65) above, the alternations are presented as **directed**, with a **basic** alternant on the left-hand side of the arrow and a **derived** alternant on the right-hand side. In coded alternations, it is generally easy to distinguish the basic from the derived alternant, because in most cases only the derived alternant involves some overt valency-changing morphology. Two more standard cases of this kind from other languages are given in (66)–(67). Mapudungun has an applicative suffix *-l*, and Chintang forms reciprocal verbs by reduplication and an affix *-ka-*.

(66) Mapudungun (Zúñiga & Lienlaf 2013)

<V.subj[1].obj[2] 2 1> → <V'.subj[1].obj[3] 2 1 3>

a. *Ti pichi malen ye-tu-y kiñe rayen colegio mew.*
 ART little girl carry-TEL-IND one flower school LOC
 'The girl carried a flower to school.'

b. *Ti pichi malen ye-l-fi kiñe rayen ñi kimeltuchefe.*
 ART little girl carry-APPL-3.OBJ one flower 3.POSS teacher
 'The girl carried a flower to the teacher.'

(67) Chintang (Schikowski et al. 2013)

<1-ERG 2-ABS V.AGT[1].OBJ[2]> → <1-ABS V'.SUBJ[1]>

a. *Cha-ŋa puchak lauri-ŋa ten-o-s-e.*
 child-ERG snake stick-ERG hit-[3SA.]3[s]P-PRF-IND.PST
 'The child has hit the snake with a stick.'

b. *Teĩ-ka-teĩ lus-i-niŋ-kha.*
 beat-RECP-beat AUX-1p[S]-NEG-BGR
 'Let's not beat each other!'

But when the alternation is uncoded, as in the Dative alternation and the Causal-noncausal alternation in English, it is often difficult or impossible to determine which of the alternants is basic. But given the setup of our database, where alternations are associated with basic verbs rather than abstract verb roots, this decision must be taken, be it in some arbitrary way.¹⁵

¹⁵ The alternative to this would have been a database with a list of verb roots (or stems), plus a list of coding frames that can be associated with the verb roots in a many-to-many fashion. Such a database would not have contained a data type "alternation" at all. We considered this alternative,

Another issue that sometimes arises with alternations is that the same alternation occurs twice in our database, but in different directions. For example, Mandinka has a Causative alternation, which turns intransitive verbs into transitive verbs, as in (68).

- (68) Mandinka (Creissels 2013)
- a. *Saatéw-o jani-ta.*
village-DEF burn-PFV.POS
'The village burned / was burned.'
- b. *Jáw-oo-lu yé saatéw-o jani.*
enemy-DEF-PL PFV.POS village-DEF burn
'The enemies burned the village.'

It also has a Noncausative alternation, which turns transitive verbs into intransitive verbs, as in (69).

- (69) Mandinka (Creissels 2013)
- a. *Kambaan-óo ye palantéer-oo teyi ber-óo la.*
boy-DEF PFV.POS window-DEF break stone-DEF OBL
'The boy broke the window with the stone.'
- b. *Kíl-oo teyí-ta.*
egg-DEF break-PFV.POS
'The egg broke.'

These two alternations are of course really the same alternation in Mandinka. The reason they are treated as two alternations for the purposes of the database is that the relevant comparison meanings happened to be 'burn (intr.)' and 'break (tr.)' and alternations had to be unidirectional for the sake of the database

13 Conclusion

Let us summarize briefly our discussion and our choices for the Valency Patterns Leipzig database (Hartmann et al. 2013). As in other areas of morphosyntax, comparison of verbal valency requires semantic comparative concepts as a basis of comparison, in particular verbal comparison meanings and microroles (§5). Distinguishing consistently between arguments and adjuncts across languages is diffi-

but decided that it would have been more difficult to extract the information that interested us from such a database.

cult, but it is not really necessary to capture the most important aspects of valency variation, because the intermediate cases (especially locational and instrumental arguments/adjuncts) show relatively little variation across languages (§4). To be comparable across languages, coding frames must primarily contain information about argument flags (cases and adpositions associated directly with the argument nominals) and about argument indexes (person forms mostly associated with the verb) (§6). It is primarily nominal arguments that are of interest for valency comparison (§7), and prominence-based splits are not encoded in the valency frame (§8). Counterpart verb forms need not be verbs in the strict sense, but can be adjectives and complex verbal expressions, but they must be fixed conventionalized expressions. Incorporated nouns that are part of such complex expressions do not count for the valency of the verbal expression (§9). Cases of apotaxis, i.e. the availability of different roles for argument expression in verbs with similar meanings in different languages, are a non-negligible difficulty for comparing valencies that must be kept in mind (§10). Alternations have to be regarded as directed in a database which asks for each verb whether it undergoes a certain alternation or not (§12).

Acknowledgements

We are grateful to Bernard Comrie and Andrej Malchukov for helpful comments on this paper.

Special abbreviations

ACT	actor inflection
AOR	aorist
BGR	background
CPL	completive aspect
D1/2/3	proximal/distal/anaphoric/demonstrative
POS	positive
PREP	preposition
UND	undergoer inflection
TEL	telic

References

- Abraham, Werner (ed.). 1978. *Valence, Semantic Case, and Grammatical Relations*. Amsterdam: Benjamins.
- Barðdal, Jóhanna. 2013. Icelandic Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary

- Anthropology. (Available online at <http://valpal.info/languages/icelandic>, Accessed on 2014-07-25)
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar, Manoj Rai, Netra P. Paudyal, Goma Banjade, Toya N. Bhatta, Martin Gaenszle, Elena Lieven, Ichchha Purna Rai, Novel Kishore Rai & Sabine Stoll. 2010. The syntax of three-argument verbs in Chintang and Belhare (Southeastern Kiranti). In Andrej Mulchokov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in Ditransitive Constructions: A Comparative Handbook*, 382–408. Berlin: Mouton de Gruyter.
- Bugaeva, Anna. 2013. Ainu Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/ainu>, Accessed on 2014-07-25)
- Campbell, Eric. 2013. Zenzontepec Chatino Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/zenzontepec-chatino>, Accessed on 2014-07-25)
- Cennamo, Michela & Claudia Fabrizio. 2013. Italian Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/italian>, Accessed on 2014-07-25)
- Chomsky, Noam A. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Comrie, Bernard & Zaira Khalilova. 2013. Bezhta Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/bezhta>, Accessed on 2014-07-25)
- Connors, Thomas J. & David Gil. 2013. Jakarta Indonesian Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/jakarta-indonesian>, Accessed on 2014-07-25)
- Creissels, Denis. 2013. Mandinka Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/mandinka>, Accessed on 2014-07-25)
- Dik, Simon C. 1997. *The Theory of Functional Grammar*. 2 vols. Berlin: Mouton de Gruyter.
- Estrada-Fernández, Zarina, Mercedes Tubino Blanco & Jesús Francisco Villalpando Quiñonez. 2013. Yaqui Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/yaqui>, Accessed on 2014-07-25)
- Goddard, Cliff. 2013. English Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/english>, Accessed on 2014-07-25)
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Grebe, Paul. 1959. *Grammatik der deutschen Gegenwartssprache*. (Duden). Mannheim: Bibliographisches Institut.
- Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. *Valency Patterns Leipzig*. Leipzig: MPI for Evolutionary Anthropology. (Available online at <http://valpal.info>)
- Hartmann, Iren, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484.

- Hartmann, Iren. 2013. Hoocak Valency Patterns In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/hoocak>, Accessed on 2014-07-25)
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntax of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries: Studies in Memory of Anna Siewierska*, 197–226. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2014. Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery* 12(2). 3–11.
- Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyacı. 2014. Coding causal-noncausal verb alternations: A form-frequency correspondence explanation. *Journal of Linguistics* 50(3). 587–625.
- Helbig, Gerhard & Wolfgang Schenkel. 1969. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: Bibliographisches Institut.
- Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
- Khurshudian, Victoria & Michael Daniel. 2013. Eastern Armenian Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/eastern-armenian>, Accessed on 2014-07-25)
- Khurshudian, Victoria & Michael Daniel. 2015. Eastern Armenian. In Andrej Malchukov & Bernard Comrie (eds.), *Valency Classes in the World's Languages*. Berlin: De Gruyter Mouton.
- Kishimoto, Hideki & Taro Kageyama. 2013. Japanese (standard) Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/japanese-standard>, Accessed on 2014-07-25)
- Koch, Peter. 1994. Verbvalenz und Metataxe im Sprachvergleich. In Werner Thielemann & Klaus Welke (eds.), *Valenztheorie – Werden und Wirkung*, 109–124. Münster: Nodus.
- Langacker, Ronald W. 1988. The nature of grammatical valence. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 91–125. Amsterdam: Benjamins.
- Lehmann, Christian. 2013. Yucatec Maya Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/yucatec-maya>, Accessed on 2014-07-25)
- Levin, Beth & Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge: Cambridge University Press.
- Malchukov, Andrej & Alexander Jahraus. 2013. Russian Valency Patterns In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/russian>, Accessed on 2014-07-25)
- Mel'čuk, Igor' Aleksandrovič. 1974. *Opyt teorii lingvističeskix modelej "smysl-tekst": semantika, sintaksis*. Moskva: Nauka.
- Michaelis, Susanne Maria & the APiCS Consortium. 2013. Experiencer constructions: headache. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, 260–263. Oxford: Oxford University Press.
- Moyse-Faurie, Claire. 2013. Xârâcùù Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/xaracuu>, Accessed on 2014-07-25)

- Nordhoff, Sebastian. 2013. Sri Lanka Malay Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/sri-lanka-malay>, Accessed on 2014-07-25)
- Peterson, David A. 2007. *Applicative Constructions*. Oxford: Oxford University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Schikowski, Robert, Balthasar Bickel & Netra Prasād Paudyāl. 2013. Chintang Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/chintang>, Accessed on 2014-07-25)
- Schultze-Berndt, Eva. 2013. Jaminjung Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/jaminjung>, Accessed on 2014-07-25)
- Seifart, Frank. 2013. Bora Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/bora>, Accessed on 2014-07-25)
- Shibatani, Masayoshi & Ketut Artawa. 2013. Balinese Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/balinese>, Accessed on 2014-07-25)
- Svartvik, Jan. 1966. *On Voice in the English Verb*. The Hague: Mouton.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Vajda, Edward J. & Elena Kryukova. 2013. Ket Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/ket>, Accessed on 2014-07-25)
- Van Valin, Robert D., Jr. & Randy J. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.
- Van Valin, Robert D., Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Vater, Heinz. 1978. On the possibility of distinguishing complements and adjuncts. In Werner Abraham. 1978. *Valence, Semantic Case, and Grammatical Relations*, 21–45. Amsterdam: Benjamins.
- Waltereit, Richard. 1998. *Metonymie und Grammatik: Kontiguitätsphänomene in der französischen Satzsemantik*. Tübingen: Niemeyer.
- Watanabe, Honore. 2013. Sliammon Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/sliammon>, Accessed on 2014-07-25)
- Zhang, Guohua. 2013. Mandarin Chinese Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/mandarin>, Accessed on 2014-07-25)
- Zúñiga, Fernando & Leonel Lienlaf. 2013. Mapudungun Valency Patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info/languages/mapudungun>, Accessed on 2014-07-25)

