

Fiat Lingua

Title: Intro to Lexical Typology

Author: Aidan Aannestad

MS Date: 07-01-2018

FL Date: 08-01-2018

FL Number: FL-000053-00

Citation: Aannestad, Aidan. 2018. "Intro to Lexical Typology." FL-000053-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 August 2018.

Copyright: © 2018 Aidan Aannestad. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Intro to Lexical Typology

or

How to avoid relexing your L1 even when your grammar is all different

Aidan Aannestad

This article is largely an attempt to break down, process and regurgitate the contents of Leonard Talmy's article 'Lexical Typologies' in *Language Typology and Syntactic Description*, volume 3. I bought this book on recommendation from the Conlanger's Library for totally different reasons, and this particular article has blown my mind so wide open that I had to sit down and rewrite its contents for myself to even begin to wrap my head around the whole new world it's presented to me.¹ I've largely kept to the structure of his article, but I've inserted comments relevant to conlanging here and there, and added some natlang examples from my own experience and reading. Hopefully this rewriting proves as helpful to others to read as it has to me to write—certainly, the information here is revolutionising my conlanging. There are huge opportunities here for extremely inventive and deeply flavourful conlanging, in ways that have to the best of my knowledge have remained mostly unexplored so far.²

Lexical typology is the study of **how languages organise their lexicons**. The whole enterprise is based on the fundamental realisation that lexicons are more than just collections of atomistic items. Words group together into categories, and those categories are often very different between languages. Categories may behave in distinctive ways when they interact with various kinds of morphology. More strikingly, words can have combinations of meanings within them, and the constituent parts of those combinations may at times be accessed directly by morphological or derivational processes, or may be simply used to group words together into categories that share patterns. As an example of the multipartite meanings words can have, compare these two English sentences:

- (1) a. I kicked the ball.
- b. I kicked Jim the ball.

These two sentences show a significant difference between their two uses of the word kick. In (1b), there's an addition of a sense of *translational motion* that's absent from (1a). The verb *kick* in the second sentence has in effect (at least) two parts—a motion and a cause by which that motion occurs. English actually also lets you separate these out, though it's less natural: *I sent Jim the ball by kicking it*. English has a huge variety of words that either contain both a cause and a motion, or contain a cause and allow the motion to be added with no additional morphology. English's method isn't the only way it can be done, though—languages can differ in how they tend to combine meanings into words. They can also differ in regard to what mechanisms they have for manipulating those combinations. Lexical typology is the study of how languages pattern in these ways.

As a warning, the nature of semantics is that almost everything is a tendency and not a rule. Lexical semantics characterises languages by how words *typically* are patterned, with no implications that a language patterns words *exclusively* in any particular way. There almost certainly are languages that have lexical categories that all pattern in one way, but it wouldn't be a surprise at all to later discover a single exception to such a pattern in that category. In fact, many (probably most) languages have at least a few lexical items

¹This might be partially a failure on the part of Talmy to be clear and/or concise, though, since I've condensed the first thirty pages of his article down to seven. Still, his goals and mine are somewhat different—he has to demonstrate where I can just assert, and I can give overviews without too many details.

²A lot of conlangers have done an excellent job of making individual lexical items unique and interesting, but I'm unaware of any conlang which has (intentionally and consciously) done things with lexical categories on a much larger scale.

from all possible lexical patterns, even patterns that no language employs widely. Plus, this being semantics, words are being repurposed and extended and manipulated all the time, and in many cases, usages that ‘should’ be nonsensical can still be supplied with creative but reasonable interpretations. Thus, everything here is *tendencies*, not *rules*, and exceptions are the norm.

Talmy’s article focuses primarily on the compositional semantics of motion and position verbs, and then moves into other, somewhat less obviously compositional properties such as aspect and causation. (He largely leaves nouns untouched.) We’ll start with looking at these properties in monomorphemic words (verb roots, basically), then move on to polymorphemic combinations of verb roots and modifiers.

1 Monomorphemic meaning complexes

Verbs of motion and position seem to have the most obviously compositional semantics of any verbs crosslinguistically, and we’ll be focusing on them for quite a while. Some of the general principles may translate into other kinds of verbs; many may not. Still, as motion verbs are often very common in a language, it’s worth examining them specifically even if nothing translates elsewhere beyond the general principle of ‘semantics can be compositional’.

With motion and position verbs, there’s a whole range of semantic referents that are relevant. I’ll list them here (non-exhaustively):

- Whether it’s motion or position in question
- What is moving or positioned (Talmy’s ‘figure’)
- The manner by which it’s moving or placed (e.g. walking, running, sticking, hanging...)
- The path any motion is moving along (e.g. towards, away from, alongside...)
- The point the path or location is relative to (e.g. towards the car, next to the building; Talmy’s ‘ground’)
- The cause that instigated the motion or location (e.g. throwing, blowing, setting, leaving...)

It’s helpful to separate these out into some basic categories. I’m modifying Talmy’s terminology a bit to make it a bit less opaque, and splitting it out into these apparently crosslinguistically basic divisions:

- MOTION (its presence or absence)
- THING (whatever’s moving or positioned)
- PATH (where the motion is along, or where the location is, uh, located at)
- PLACE (the reference point the path is relative to)
- CO-EVENT (a combination of manner, cause, and other similar things)

Different languages combine these into verb roots in different ways—a phenomenon Talmy terms ‘conflation’. As mentioned above, many languages have lexemes from a variety of patterns; but typically languages have one pattern that they prefer,³ which largely means that it’s the pattern with the most lexemes and/or it’s the most natural or colloquial pattern. Let’s examine a familiar pattern first.

³Other languages have split systems, where pattern preference is per semantic category rather than a whole-language setting. As far as I know, no language lacks preferences entirely.

1.1 MOTION + CO-EVENT

This is English's preferred pattern. Like I mentioned above, English has a huge variety of verbs that conflate motion and co-events of various kinds. Here are some examples:

- (2) a. The book lay on the table.
- b. The rock sank into the depths.
- c. I left my book on the table.
- d. The ship shot a cannonball at the fort.

We can see several combinations in just these sentences. (2a) conflates location and manner, (2b) motion and manner, (2c) location and cause, and (2d) motion and cause. English allows most, possibly all (colloquially, at least), verbs that could conceivably be used as a co-event to a motion to gain a motion component 'for free', without any morphological change—e.g. *I texted him that link*.

There are many ways in which English further extends this pattern:

- The motion or position sense may be less basic than just 'be located' or 'move': the above example *I texted him that link* incorporates 'give' more than just 'move', and *The snow blanketed the lawn* incorporates 'cover' more than just 'be located'.
- There may be causal complexity to the motion component—in e.g. *I scared him away*, the motion component is not just 'move', but 'cause to move'.
- The motion may be entirely metaphorical! In *I choked him to death*, the movement component isn't translational through space, but instead through different states.
- The relation between the co-event and the motion may be one (or possibly more) of a number of types. In *She wore a green dress to the party*, the co-event is simply an accompaniment to the motion—it's not really a manner. In *The glass broke all over the table*, the breaking doesn't exactly cause the movement onto the table, it mostly just precedes it.

English is not the only language that does this—it seems fairly widespread throughout the world (though Talmy doesn't mention anything regarding how widespread this means of deriving a motion + co-event from a non-motion verb is). Talmy explicitly gives examples as varied as non-Romance IE, Finno-Ugric, Mandarin, Ojibwe and Warlpiri.

As a conlanger, there are a few ways to get variations inside a system like this. You can pick which sets of verbs allow the motion to be added 'for free' and which require some sort of derivational morpheme, and you can pick which verbs allow which kinds of extensions from the list above (and which allow them only with derivational morphemes). You can also restrict which kinds of co-events—causations, manners, etc—which verbs allow, and maybe just blanket restrict some from occurring—I can easily see a system in which a phrase like *The glass broke all over the table* is disallowed, because the system doesn't allow a preceding co-event to be conflated together with the motion that follows it.

1.2 MOTION + PATH

The next major preference (and apparently the most crosslinguistically widespread) is for conflating motion with the path it's moving along. Spanish does this: 'He left the school' is *El salió de la escuela*, but there's nothing structurally comparable to English 'He ran out of the school'—all you can say is *El salió de la escuela corriendo* 'He left the school running', which 1) requires an additional, separate verb and 2) isn't all that natural under most circumstances.

There is a smallish number of basic path categories, which Talmy lists as the following:

- (Be) at a point
- To a point

- From a point
- Via a point
- Along an extent
 - Toward a point
 - Away from a point
 - Along an extent ending at a point
 - Along an extent starting at a point
- Across a distance
 - From a point to another point

The idea of ‘path’ can be further augmented with what Talmy calls the ‘conformation’, which provides information about the situation of the point being referenced. For example, the English prepositional phrase *in the thing* can be broken down into ‘at a point which is located inside of the thing’; *onto the thing* is ‘to a point which is located on the surface of the thing’; *past the thing* is ‘via a point which is to the side of the thing’. These complex path constructions can be lexicalised wholesale into verbs: Spanish *salir* is ‘move from a point which is inside of some thing’. I imagine there’s a wide range of possible conformations a language could choose to care about—interiors and surfaces seem most common, but others are readily available.

Additionally, paths can be specified with a deictic ‘to the speaker’ or ‘away from the speaker’ status, commonly seen in verbs such as English *come* and *go*. It seems like verbs don’t tend to have both a basic path component and a deictic component in a single morpheme—it seems a bit cumbersome to take the range of possible basic path verbs and multiply it by three to get ‘to me’, ‘from me’ and ‘unspecified relation to me’ verbs for every one of the language’s path types. One avenue for interesting less-natlangy conlanging could be to increase the number of deictic categories allowed, and have things like ‘in front of me / across the front of me’, or ‘to / along my left side’, or similar things.

In Spanish, the main verb in a sentence is the motion + path verb and the co-event is a separate non-finite verb (eg the above *salí corriendo* ‘exited running’). Japanese, which also prefers motion + path conflation, seems to have the opposite pattern: the main verb is the co-event, and the motion + path is the non-finite form:

- (3) gakkou=wo de-te hashit-ta
 school=OBJ exit-CVB run-PAST
 ‘[He] ran out of the school’; literally ‘[He] ran, exiting the school’

The above example is also perfectly natural and colloquial, unlike its Spanish counterpart. In fact, Japanese prefers making a motion + deictic path verb the main verb when such a thing is relevant, and leaving both the manner and the rest of the path information in non-finite verbs:

- (4) gakkou=wo de-te hashit-te ki-ta
 school=OBJ exit-CVB run-CVB come-PAST
 ‘[He] came running out of the school’; literally ‘[He] came, running, exiting the school’

You can tell from the example glosses involving words like *exit* that English has a set of motion + path verbs besides its motion + co-event verbs, and you might be wondering why English isn’t also in this category. It’s because these verbs in English are markedly literary or otherwise unusual, and they’re not English’s preferred (most natural and colloquial) means of talking about motion events. Plus, they’re all loanwords from Romance languages or from Latin via a Romance lens, and thus they’re part of a separate system that’s been loaned into English on top of its existing native system.

1.3 MOTION + THING

This is the most crosslinguistically unusual among the patterns that a language might overall prefer, and it seems to be widespread only in (Western?) North America. This pattern involves conflating motion with the thing it is that's moving. Talmy gives examples from Atsugewi, a (recently dead) language from northern California that he's done fieldwork in, which has verb roots like the following:

- *-lup-* 'for a small, polished, round object (eg a round candy or a hailstone) to move / be located'
- *-t'* 'for a smallish, flat object that could be fixed onto something (eg a stamp or a button) to move / be located'
- *-t/faq-* 'for a slimy lumpy object (a toad, a cow dropping) to move / be located'
- *-swal-* 'for a limp, linear object suspended by one end (eg a windsock) to move / be located'
- *-qput-* 'for loose and dry dirt to move / be located'
- *-st'aq'* 'for runny, unpleasant material (eg mud, entrails, rotten tomatoes) to move / be located'

Exactly what kind of, say, runny unpleasant material can be further specified by supplying a full noun phrase as the verb's argument. In Atsugewi, these verbs can be extended without morphology to have causative meanings as well (e.g. 'cause the runny and unpleasant material to move').

This kind of conflation of motion with the thing that's moving is what underlies the frequently-raised example of Navajo giving verbs. It's not that Navajo is somehow deeply, incomprehensibly alien; it's just that Navajo has conflated giving + the thing being given in exactly the same way English conflates giving + the method of giving (in e.g. *kick him the ball*, *text him the picture*; phrases which themselves cannot easily be translated into Navajo).

These sorts of thing-conflation verbs seem to be based in the same kinds of semantic categorisations that other kinds of noun classification mechanisms (such as East Asian counter classifiers) tend to use. These seem to focus primarily on the shape (and/or position) of solid objects or the consistency of materials that don't have a shape, with secondary properties such as texture or use or other cultural evaluation (such as 'unpleasant' with Atsugewi *-st'aq'*)—compare the above examples with Japanese *sanmai* 'three wide, flat sheets' and *sanbon* 'three long, thin objects'. One way to make a very alien conlang would be to make a similar noun classification system based on some other set of properties that are less salient to humans, such as colour or composition. You could also expand these classifications into semantic spaces that would have been much less interesting or accessible to humans historically: you could have verb roots meaning things like like 'for a rideable, self-propelled vehicle to move' or 'for digital information to be located' or so on.

Notice in the above Atsugewi example that the verb roots themselves don't make a distinction as to whether actual motion is involved. We'll talk about these sorts of patterns later.

1.4 Less common patterns

The above three patterns are found as the overarching preference in some languages. There are other conceivable patterns that are found in niche cases, but never as a language-wide preference. As a conlanger, these unusual patterns provide an excellent (and extremely underutilised, I think) way to make a language that sees the world in a well-and-truly alien way—take one of these to be your language's overall preference, and you'll have something that is really visibly and almost viscerally outside the realm of natural languages.

1.4.1 MOTION + PLACE

This is possibly the most alien, at least from a statistical standpoint; Talmy mentions that he's unaware of any language that prefers this conflation even in a subcategory of its verbs. This pattern involves conflating motion with the point or extent the motion is relative to. English has a few words that possibly exemplify this, like *deplane*, where the root *-plane* could be seen as meaning 'move relative to an airplane', or *embark*

with *-bark* seen as ‘move relative to a ship or boat or similar vehicle’; but those verbs aren’t part of productive patterns (e.g. there’s no *transplane* for ‘move through an airplane’ or *circumbark* for ‘move around a ship’). Talmy speculates that this could be a fully possible preference for a natlang, but simply so rare that there aren’t enough languages in the world to make it statistically likely to appear.

1.4.2 MOTION + more than one conflated component

This one seems to be avoided on the basis of trying to keep a manageable lexicon size. English has a few motion + path + place verbs like *box* (‘move into a box’) and *shelve* (‘move onto a shelf’), and motion + path + thing verbs like *powder* (‘move powder onto something’), but these are relatively uncommon. Japanese giving verbs seem to conflate giving + deictic path + relative social status of participants—*ageru* ‘give from speaker to someone higher’, *kureru* ‘give to speaker from someone higher’, *kudasaru* ‘give to speaker from someone much higher’—though I’m finding it hard to find verbs where the path and status coincide in a way that indicates the speaker themselves to be someone higher. A language with this as its overall preferred strategy would risk having just too much unanalysable complexity—it’s the same reason that suppletion for grammatical things like tense or negation is relatively limited in its application.

1.4.3 MOTION + nothing else at all

In this pattern, you have one verb ‘be located’ and one verb ‘move’ (and maybe one verb ‘cause to move’), and all the other information has to be supplied in some other way. Spanish has this with location verbs—all you have is *estar*, with no information about the cause or the location or the thing located. Many, maybe most, languages have a verb or two that only express motion with no additional components (eg English *move*), but this doesn’t seem to be favoured as a preferred strategy. Talmy suggests that this is because such a verb root ‘move’ would have to be restated every time there’s a motion event, even when the motion is inescapably implied by the presence of some other component of the motion—e.g. if you say ‘move into a box’, it’s clear from the path component in ‘into’ that there must be motion, and forcing speakers to say ‘move’ outright every time anyway is just inefficient. You might as well piggyback *something* off of it.

1.4.4 MOTION + a minimum of other information

Some languages do seem to have something approaching the above as major patterns. In these cases, there’s some other component present in the verb root beyond just motion, but there’s a very small number of distinctions that can be made by verb roots. Talmy mentions Southwest Pomo, which conflates motion + thing, but the only distinctions made in the verb root are the plurality of the thing—there’s three roots for motion: *-w* ‘for one thing to move’, *-?da* ‘for two or three things to move’, and *-p^hil* ‘for several things to move together’, and those roots appear in most kinds of motion events. Japanese conflates location + thing in regard to animacy only, with a contrast between *aru* ‘for an inanimate thing to be located’ and *iru/oru* ‘for an animate thing to be located’. Similarly, Hindi non-agentive (non-‘intentional’, I assume) motion verbs conflate motion + only the deictic portion of path, and thus non-agentive motion events are built around just the two roots ‘come’ and ‘go’.

1.4.5 Split and complementary systems

Some languages lack a clear overall pattern preference, because different classes of verbs have different preferences. Emai, a language from Nigeria, apparently prefers path conflation with self-agentive verbs (in cases of things like ‘walking’, where the agent causing the movement and the thing moving are the same), and prefers co-event conflation with anything else. Spanish, like we saw above, prefers path conflation (mostly) with movement verbs and zero conflation with location. Tzeltal has a wide range of thing conflation roots, but it only uses them in situations where the thing in question is or ends up physically supported by something else.

Some languages may have two patterns that are mostly equally preferred. Greek, apparently, is quite happy to use either a path verb plus a non-finite co-event or a co-event verb plus a preposition for the

path, with neither being markedly less colloquial or natural. In this case, what’s different is the degree of backgrounding of the various components—something we’ll get to later.

No language seems to have verb classes with a fully random assortment of patterns—individual verb classes, at least, seem to almost always have a characteristic pattern, even if there are many exceptions. Latin seems to be a close approach to such a random distribution, because it’s in the process of transitioning from an IE-style co-event conflation preference to a later Romance path conflation preference, but even then, no verb class has a total lack of a clear pattern preference.

1.5 Lexical aspect

Lexical aspect is a bit of a hard thing to define exactly—Talmy defines it as ‘the distribution of an action through time’, i.e. the structure of an event over the time it occurs in; which seems to cover most of the relevant parts of lexical aspect. We’ll talk about a couple of others here as well.

Talmy lists six different kinds of state transitions, which are a bit different from the four canonical main types of lexical aspects (‘state’, ‘activity’, ‘accomplishment’ and ‘achievement’). I’m going to combine the two systems together in this chart.

- One-way transitions
 - Lengthy one-way transitions, such as *grow up*
 - Instantaneous one-way transitions, such as *explode*
- One-way resettable transitions
 - Lengthy one-way resettable transitions, such as *fall*
 - Instantaneous one-way resettable transitions, such as *wake up*
- Full, single-cycle on-and-off transitions, such as *flash*
- Multi-cycle on-and-off transitions, such as *breathe*
- Unbounded states and activities
 - States that will continue if left alone, such as *love*
 - Activities that end without a continuous input of effort, such as *sing*
- Gradients, such as *widen*

Each of these behaves differently in regards to things like certain adverbs (you can *widen more* but you can’t *wake up more*), as well as with derivational or grammatical aspect morphemes. As far as I’m aware, these are universal semantic categories—languages have some verbs in every category, and ways to move verbs between categories (likely between every pair of categories). Still, there’s room for a lot of variation: languages differ as to which category certain kinds of things have their most basic root placed into. For example, do you have a transition verb ‘die’, or a state verb ‘be dead’? As with other things, semantic categories of words are often treated together—English, for example, seems to treat short-term, easily-changed states as activities, such as with *sit*. Words can often fit into multiple categories depending on their context—for example, English *look at* can be either an activity (‘looked at the picture for five minutes’) or a one-way resettable transition (‘looked back and forth at the two pictures’). Notice that *watch* is only a synonym for the activity sense of *look at*, and not the transition sense.

Which category you choose to put a given group of words into is very relevant for conlanging, but for our discussion, these aspects will become much more relevant when we get to multimorphemic verb complexes. It’s worth noting, though, that languages that have productive and regular ways of putting verb roots into one of these categories often have very few verb roots that are inherently in that category. For example, a language with a morpheme that creates multi-cycle transitions will likely treat words like ‘breathe’ as single-cycle transition verbs, and will mostly access the multi-cycle version only through adding morphology.

1.6 Causation

Verb roots can differ on whether the event they're describing involves a cause or not. For example, *the table broke* implies no cause whatsoever, but *I broke the table* has incorporated a causing event (even if an unspecified one).

Talmy identifies nine separate types of causation-related event, using English *break* as the example verb when it's valid (which is almost all of them!, though not always naturally):

- Autonomous, independent event—*the vase broke*
- Resultant event—*the vase broke* (because a ball rolled into it)
- Causing event—*the ball's rolling into the vase broke it*
- Instrument causation—*the ball broke the vase* (by rolling into it)
- 'Author' causation—*I broke the vase* (by accidentally rolling a ball into it)
- 'Agent' causation—*I broke the vase* (by quite intentionally rolling a ball into it)
- Undergoer event—*I broke my arm* (when I fell)—not truly causative, but still patterns with causatives sometimes
- Self-agentive causation—*I walked to the store*
- Inducive causation—*I sent him to the store*

Different verbs in different languages may be valid for different subsets of these. Talmy gives the English example triad of *die/kill/murder*, where *die* is required for autonomous and resultant event, *kill* is required for causing event, instrument causation and author causation, and *murder* is required for agent causation. Other sets may overlap—*smash* is valid for many of the senses of *break*, but not for autonomous event (**the vase smashed*) and questionably for author causation (*?I accidentally smashed the vase*).

Verbs may also be able to jump between these categories morphologically. A conlang I'm working on at the moment has an affix that takes an autonomous verb or an author verb and turns it into an agent or self-agent verb: *xagu* 'go bad, get moldy' → *xaguvi* 'leave to ferment'; *máro* 'die' → *márovi* 'kill oneself, engineer one's own death'. Japanese has semi-predictable lexical pairs for many verbs on this spectrum: *kowareru* 'break' (autonomous/resultant event) vs *kowasu* 'break' (causing-event and beyond); *tsuku* 'get stuck on' (autonomous/resultant event) vs *tsukeru* 'stick on' (causing-event and beyond). In other cases, it's constructions rather than single morphemes that switch between types—*He drove to the store* → *I convinced him to drive to the store* (self-agentive → inducive); Spanish *abrió la puerta* 'he opened the door' → *la puerta se abrió* 'the door opened' (lit. 'the door opened itself') (agent → autonomous). Some verbs are simply unavailable or highly marked with some types—*?I mislaid myself somewhere* (self-agentive); *?I convinced him to mislay his pen* (inducive).

Some languages treat the undergoer-event kind of causation with special syntax. Spanish, for example, places the affected party as something like an indirect object with a passive verb: *se me quebró el brazo*, literally 'the arm was broken (to) me'. Barai, a language from Papua New Guinea, treats undergoers morphologically as objects (eg 'my head hurts me' for 'I have a headache'), but syntactically they behave as if they were subjects for things like clitic placement and cross-clause coreference.

1.7 Aspect and causation interacting

Causation is often invalid with state readings, and is frequently tied up in state-change situations. Languages often group things together in this zone. Talmy identifies a sort of staged gradation from being in a state through autonomously changing to causing change. Various languages, and more often categories of verbs within languages, will choose a part of this scale as 'basic' and use morphology or syntax to access the rest of it. Japanese verbs regarding postures (standing, sitting) tend to pick autonomous change as their most basic type, and access the rest indirectly:

- (5) a. boku=wa tat-ta
 1sg=TOP stand-PAST
 ‘I stood up’ (entering the state)
- b. boku=wa tat-te i-ta
 1sg=TOP stand-PERF-PAST
 ‘I was standing’ (being in the state)
- c. hon=wo tate-ta
 book=OBJ make.stand-PAST
 ‘[I] stood the book up’ (putting into the state by agent causation)⁴
- d. kodomo=wo tat-ase-ta
 child=OBJ stand-CAUS-PAST
 ‘[I] stood the child up’ (putting into the state by inductive causation)

Spanish chooses the put-into-state form as the basic form of posture verbs: *acostar* ‘lay⁵ (someone) down’, *acostarse* ‘lay down’ (lit. ‘lay oneself down’), *estar acostado* ‘be laying down’ (lit. ‘be having been laid down’). English (largely) picks the be-in-state option, clearer with past: *I stood (there)*, *I stood up*, *I made him stand up*. Notice how in English, the put-into form is accessed via the get-into form (*stand* → *stand up* → *make stand up*)—to go from one end to the other, you have to go via the middle. In Spanish, there’s no such requirement—the be-in form is accessed independently from the basic put-into form, without having to go through get-into. German has a different pathway—it goes directly from *sitzen* ‘sit’ (be-in) to *setzen* ‘make sit’ (put-into), and then back from there to *sich setzen* ‘sit down’ (get-into).

Modern Standard Arabic posture verbs group be-in and get-into together, and treats the combination as basic: *nāma* can mean either ‘he was laying down’ or ‘he lay down’. Some English state change words merge get-into and put-into: *The water was frozen*, but *The water froze / I froze the water*. English has a small class of words including *hide* which merge all three together—*I hid there for an hour / I hid there when I heard the noise / I hid him there*. Talmy describes Atsugewi posture verbs as having no basic choice and requiring morphology of some kind in all three cases; though that morphology tends to group get-into and put-into together.

Languages vary on how pervasive patterns of the above tend to be throughout the language. Patterns seem to hold more for semantic classes of verbs rather than whole languages, and even within classes, having non-conforming verb roots is not all that surprising.

Interestingly, while *entering* states seems to be quite a core concern of language, *leaving* states is typically an afterthought. The same verb root is never available for both a be-in/get-into/put-into reading and a get-out-of reading, even with morphology: **He hid out of the attic / *He unhid from the attic*. Additionally, there seems to very rarely be a morphological pattern for leaving a state that is as core as the above-discussed entering morphology. These sorts of things tend to be expressed through multiple stages of derivation—such as English *decentralise*, which is the get-into/put-into *centralise* plus a reversal morpheme, not just *central* with one state-exit morpheme; or Japanese *shirokunaku naru* ‘stop being white’, which is literally ‘become not white’.⁶ Having even just one root verb for ‘exit a state’ comparable to *become* for ‘enter a state’ is rare if it occurs at all. Again, here’s another opportunity for subtly but pervasively alien conlanging—elevate state departure to the same status as state entry, or even reverse their positions and make state entry about un-leaving a state!

⁴The causation morpheme *-e-* here is basically fossilised and not productive.

⁵I’m not drawing English’s *lay/lie* distinction in the example glosses, because I don’t have that distinction in my own speech and I don’t remember which forms are which. Don’t read anything into my choices of forms.

⁶Japanese has actually quite cleverly taken advantage of a historical accident here, in that the standard Japanese negation marker is a grammaticalised adjective and can still be inflected like an adjective. Other Japanese dialects with negation markers not from adjectives can’t do this. With verbs, this often implies a permanence - *utawanaku naru* ‘become not singing’ → ‘stop singing’ doesn’t just mean ‘end a concert’ or ‘stop partway through a song’, but ‘cease from singing with no intent / ability to sing ever again’.

1.8 Personation

Some verbs can vary as to whether they're considered to primarily involve an agent acting on themselves or on some other person. For example, French *Je me raserai* is literally 'I shaved myself', not just 'I shaved'; and **Je raserai* without a reflexive is invalid. English has a range of verbs that all by default imply action on oneself, including things like *I bathed*, *I showered*, *I undressed*; all of these require reflexive morphology in French. These verbs in English are largely valid for both kinds of personation, though—*I shaved him*, *I bathed him*, etc., are all valid (though **I showered him* isn't, at least for that meaning). This is not necessarily the case for self-action verbs like this—Atsugewi has a class of self-action-only verbs which require a benefactive affix to allow for e.g. bathing someone else.

Note that this isn't necessarily correlated with the verb's transitivity. English *get* is a similar kind of self-action verb (in the sense 'obtain for oneself'), and has to appear with a benefactive phrase *get for someone* before it can affect someone else.

1.9 Valence

'Valence' refers to the number of arguments a verb takes (subject, object, etc.), and what semantic roles each argument has. The distinction between transitive and intransitive verbs is the simplest and most famous valence distinction, but there are other sorts of valence configurations a verb can take. Traditionally, 'valence' has referred purely to the number of arguments, but often it is just as interesting to talk about which semantic role is assigned to each argument, and Talmy extends 'valence' to refer to the argument structure's organisation in general.

Some verbs can vary across languages as to how many arguments they take. A good example is weather verbs, such as *it's raining*. In English, these largely take the dummy subject *it*, which cannot be replaced by anything else, and are in a way 'zero valence'—you can't, for example, say **water is raining*. In Japanese, however, *furu* 'fall as precipitation' has a slot for a subject, and 'it's raining' is expressed as *ame=ga futte iru* 'rain is falling as precipitation'.

It's also possible to have words with the same valence, but a different organisation of arguments, and thus a different perspective on the event they're describing. Consider the following English examples:

- (6) a. I stole his money off him.
- b. *I stole him of his money.
- c. I robbed him of his money.
- d. *I robbed his money off him.

Both *steal* and *rob* describe the same event, namely, the taking of property from an unwilling donor, but they differ as to which non-agent participant they take as the object and which as an oblique phrase. Many languages have pairs of words like this. Some words, such as English *load*, may have more than one valid perspective—you can *load a truck with stuff* just as well as you can *load stuff onto a truck*.

2 Multimorphemic meaning complexes

The thing that most expanded my mind about this paper was Talmy's concept of *verb satellites*—morphemes that can be added to a verb root to add to or specify the verb's meaning further. These can take any form—affixes, other verb roots in serialisation, adpositions used adverbially⁷ and other particles, compounded or incorporated verb or noun roots, and so on;⁸ and languages typically employ a variety of forms for satellites, frequently even within single semantic spaces. The unifying concept is that they combine with the verb root to create a single verb complex. For example, take the English sentence *would you go run take back down*

⁷I suspect that in English at least, the category of 'prepositions without objects' is just a class of verb satellite particles, not actually a class of adverbs.

⁸It's not clear from Talmy's description whether he considers multi-part periphrastic constructions like English *be V-ed* or Japanese *V-te iru* as satellites, but I see no reason to exclude them.

that sign?, where *go run take back down* is a verb complex comprised of the root *take* plus four satellites. *Go* here is used to indicate that the subject must move to some other place to do the action; *run* means that the action should start as soon as possible and likely will finish not long after, *back* indicates that the action is returning something to a previous state, and *down* is a path satellite indicating that the taking motion occurs along a downward trajectory. *Go* and *run* are verbs in serialisation, and *back* and *down* are simple particles.

As the above example shows, there's quite a wide range of meanings that can be incorporated into a verb via satellites. You can have any of the motion components that the main verb lacks: path, thing, place, or co-events of various kinds. English uses its preposition-derived satellites for path usages quite frequently—*walk past*, *run away*, etc. Caddo, a language from northeast Texas, uses what used to be incorporated nouns as a kind of thing satellite affix: *widiŋ dânnida?ka* ‘he found (‘powder-found’) some salt’, but *tsútsu?kanida?kah* ‘he found (‘liquid-found’) some milk’. Atsugewi has not just co-event affixes, but separate classes of affixes for manner and cause.⁹ You can also add further specification about a component the verb already has—Japanese has words like *hairikomu*, which is the path verb *hairu* ‘enter’ plus the path satellite *komu* ‘get inside and stay inside’—meaning in combination ‘go in with the intention of staying a while’.

Satellites also can be used to alter a verb's lexical aspect or causation status. English has phrases like *eat up*, where *up* turns the activity verb *eat* into a one-way transition verb with a clear endpoint. Japanese has things like *yaritsuzukeru* ‘keep doing’, or *yarikakeru* ‘get partway through doing’. These sorts of satellites typically carry a bunch of additional semantics along with the simple aspect or causation change—see the above English example with *go run take*, where the satellites seem to be aspectual somehow, but in very non-basic and elaborated ways. These satellites can be extremely polysemous at times, and may have several apparently unrelated meanings depending on the kind of verb they're being used with—English *up* can be an aspect modifier like with *eat up* or a path addition with verbs like *pull up* (or ambiguous for either, in the case of *walk up!*).

Talmy also has a category of what he calls ‘satellite prepositions’, meaning words that act simultaneously like verb satellites and like prepositions that take objects, as in English *I drove past him* (c.f. *I drove past*); Mandarin also appears to have a similar class of words. I suspect this category should be expanded to include any sort of satellite that also modifies the verb's valence, which would then also include things like applicatives that can be used as satellites while still retaining their valence-change effect. My conlang Emihtazuu makes extensive use of applicative satellites that also add objects to the verb—*li lánikósa* ‘[walk away from] [the person]’, not **lízi lánikósa* ‘[walk away] [from the person]’.

Languages vary quite widely on how much information it's possible to supply by satellites. Spanish seems to have very little indeed possible, and may in fact lack satellites entirely. Nearer the other end of the spectrum is Atsugewi, which has words like this:

- (7) [m²ast'aq'ipsnuk:a]
 ?w- ma- st'aq' -ipsnu -ik -a
 3sg.FACT- by.feet- runny.material.move -inside.volume -toward.me -3sg.FACT
 ‘he moved runny unpleasant material inside a three-dimensional enclosure towards where I am by using his feet’ (→ ‘he tracked mud into the house’)

And that's only a verb and three satellites! English can do at least one more, as in the above *go run take back down*.

Japanese seems to be in an odd situation with regards to the syntactic status of satellites, as it seems (by my analysis) that the semantically ‘core’ verb root almost always appears in some sort of non-finite or otherwise dependent form, while the syntactic head of the construction is the last of a group of satellites (here shown by being the only one that takes tense):

- (8) a. maze -komi -tsuzuke -te iru
 mix -enter.and.stay.in -continue -PROG

⁹Interestingly, these still seem to be largely following Atsugewi's preference for thing-oriented perspectives. An example Talmy gives is a cause affix meaning ‘by poking with a long, thin object like a stick’.

‘[subj] is keeping on mixing it in’

- b. maze -komi -tsuzuke -te i -ta
mix -enter.and.stay.in -continue -PROG -PAST

‘[subj] was keeping on mixing it in’

Compare this with its above strategy for adding full co-event verbs, where the path-conflated ‘main’ verb is also demoted to a non-finite form and the added co-event verb becomes the head verb. In short, while we might expect that the term ‘satellite’ would imply a syntactically dependent status as well as a semantically dependent status, Japanese seems to dissociate the semantic status of ‘core verb root’ from the syntactic status of construction head. I’ve seen some examples of Papuan languages that appear to do the same thing, with verb serialisation constructions where the inflected (or most fully inflected) verb is the added satellite rather than the semantically main verb. This is in contrast to e.g. Spanish, where added co-event verbs (since satellites seem to basically nonexistent in Spanish) are the dependent and non-finite forms. English has a bit of an odd relationship with this status—when there’s a serialisation construction with multiple verbs, such as *go run take*, it’s not at all clear which if any is the head, and thus there’s no clear way to inflect the complex—is its past tense *went ran took?* *Go run took?* My mental grammar even produces attempts like *go ran take*. None of the above is acceptable, though, and there’s no right answer—the least bad seems to be *went and ran and took*, which is very cumbersome and not even a serialisation construction anymore! I suspect that this is because English serial verbs are still quite new to the language, and thus, while serialising bare verb stems is now fine, when you try and inflect those stems, it still sounds odd.

Notice as well that there seems to be a hierarchy of satellite placement in the above examples, where motion components are placed closest to the verb, then the more lexical aspect, then the more grammatical aspect. This hierarchy seems intuitively reasonable; I haven’t checked it out in other languages. It would fit well, though, with theories of modifier ‘layeredness’ such as that found in Role and Reference Grammar, adding to or possibly mixing into the innermost series of layers. Motion component satellites are clearly much more interior than, say, tense or illocutionary force markers.

3 Backgrounding and foregrounding

A lot of the considerations involving lexical conflation and satellites and so on revolves around the question of *backgrounded* versus *foregrounded* information—to some degree, this opposition is what I’ve been talking about the whole time. Backgrounded information can be supplied without distracting the hearer from the main point of the sentence; foregrounded information is interpreted as being part of what the speaker considers especially important—the speaker’s motivation for saying that sentence may be primarily to get that information across. It’s a bit difficult to construct an example out of context, but Talmy gives this pair of sentences (which I modified slightly for colloquialness):

- (9) a. I flew to Hawai‘i last month.
b. I took a plane to Hawai‘i last month.

In (9b), the use of the plane is much more salient, and seems likely to be in contrast to some other possible method of travel—at the least, you probably wouldn’t phrase it this way if the plane itself wasn’t somehow relevant to the point you’re about to make. In (9a), the fact that the going to Hawai‘i was specifically by plane is unimportant, and the contrastive reading, while possible, is much more of a stretch.

Conflation and satellites are strategies that languages use to put more information into a sentence without distracting from the sentence’s foreground. Information conflated into a verb root, and information supplied by a satellite, is almost universally background, and languages often have alternative ways to phrase such information in order to bring it to the foreground. Sometimes, you have language-specific restrictions on what or how much can be backgrounded: compare the English sentence *the man ran back down into the cellar* with the four possible Spanish rephrasings:

- (10) a. *El hombre corrió al sótano* ‘the man ran to the cellar’

- b. *El hombre volvió al sótano corriendo* ‘the man returned to the cellar running’
- c. *El hombre bajó al sótano corriendo* ‘the man went down into the cellar running’
- d. *El hombre entró al sótano corriendo* ‘the man entered the cellar running’

Not one of the above sentences captures all of the content in the English sentence, and in the last three, the running is specifically foregrounded in a way it isn’t in English. Spanish has no mechanism to allow it to say anything equivalent to *the man ran back down into the cellar* with the same level of compactness, and trying for any compactness at all requires altering the relative importance of the various pieces of information. Similarly, English cannot translate the Atsugewi word *mʔastʔaqʔipsnuk:a* in (7) above without having an absurdly cumbersome amount of foregrounding, especially relative to how non-specific most of it is (‘he moved runny material inside a three-dimensional enclosure towards where I am by using his feet’). Atsugewi has no more satellites here than English would allow, but they refer to entirely different kinds of information from what English has satellites for. This is part of what makes languages like Atsugewi feel so different to English speakers—the information they allow in the background is extremely different from the information languages like English allow.

In some ways, it feels to me like a good deal of the ‘flavour’ of a language, or its ‘perspective’, comes from what it allows in the background. As I said, Atsugewi feels very different from English, on a very deep and fundamental level. English in turn feels very different from Spanish. And you can go much farther—Japanese and Korean allow you to add as background information the relative social status of not only the referents but also the speaker and listener of a sentence,¹⁰ which gives its own strikingly unique flavour. I can imagine other languages with other concerns—whether an action is for someone’s benefit or detriment, or whether the actor or speaker gains or loses by an action; what time of day or year an action is performed at; what social conditions an action is performed under (eg shameful compulsion, permission granted as an honour, etc; cf Japanese’s satellite *-yagaru* ‘do without the social status such an action requires; by doing imply a claim of higher social status than deserved’); and less-naturalistic or less human languages would likely have much more odd and unfamiliar pieces of information backgroundable.

This is a major way in which cultural worldview and language can interact; and while that interaction has been greatly overstated at times,¹¹ the fact that a language allows, say, relative social status to be incorporated and backgrounded does say something about the importance of relative social status to the speakers’ culture. I would love to see conlangers get very creative with backgrounded categories of information—what sorts of things would alien cultures find important enough to background? What sorts of things might human cultures conceivably find important enough to background? And don’t get distracted by the cultural possibilities and forget the largely acultural things—‘path of motion’ is hardly a very cultural consideration, for example! There’s such a wealth of possibility here, and I’m rather disappointed by how little I’ve seen this area explored in conlangs. That’s a lot of why I’ve continued writing this article, even though by now I think I’ve fairly well understood the concepts Talmy was trying to get across—I’m hoping to get those across to other people, and encourage them to experiment with these kinds of things. I hope I’ve done a good and understandable job of doing so, and I encourage conlangers to be just as creative with their lexicon as they are with their grammar!

¹⁰Though the speaker/listener honorific information is more just a grammatical affix rather than a satellite.

¹¹Edward Sapir’s fieldwork on western North American languages similar to Atsugewi was what inspired the formulation of the Sapir-Whorf hypothesis—he saw the vastly different backgrounding possibilities and concluded that they stemmed from a fundamentally different conception of the world.