

Fiat Lingua

Title: Against Word Order & Alignment

Author: Logan Kearsley

MS Date: 05-30-2023

FL Date: 11-01-2023

FL Number: FL-000092-00

Citation: Kearsley, Logan. 2023. "Against Word Order & Alignment." FL-000092-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 November 2023.

Copyright: © 2023 Logan Kearsley. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



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

Against Word Order & Alignment

as categories guiding the design of languages

Logan R. Kearsley

What order do you put your subject, verb, and object in, and how do you tell the subject and object apart?

These two questions introduce the typological concepts of *basic word order* and *morphosyntactic alignment*, and in each case it is easy to create a finite list of possible arrangements. There are six ways to arrange a list of three distinct items, so *obviously* there are six “basic word orders” for simple transitive sentences:

VSO SVO SOV VOS OVS OSV

Similarly, there are five ways to divide up three elements (intransitive S and transitive A and P) into complementary subsets, which gives us 5 possible morphosyntactic alignments:

Nominative-Accusative:	(S A) (O)	
Ergative-Absolutive:	(A) (S O)	
Transitive:	(S) (A O)	(AKA “Monster Raving Loony Alignment”)
Direct:	(S A O)	(AKA “neutral alignment”)
Tripartite:	(S) (A) (O)	

For a beginning conlanger, these are useful, easy-to-understand categorization systems that can help with exploring structures different from one’s native language. Want to make sure your language Isn’t Like English? Just pick one of those basic word order options that *isn’t* SVO, and stick with it! Or pick an alignment that *isn’t* nominative-accusative, and stick with it! Almost everyone has, at some point, created an ergative-absolutive language for just this reason—because it’s *different*, and *cool*.

But despite being *obviously* exhaustive, examining natural languages will eventually reveal that these categorizations are, actually... *not* exhaustive. Because, despite appearing obvious, they rest on some assumptions which do not actually hold in every language! For example, in *A Grammar of Chukchi*¹, Michael Dunn states that

It probably doesn’t make descriptive sense to claim that Chukchi has some kind of basic, syntactically defined word order (i.e. a word order typology as discussed by Greenberg 1963 and many others).

¹ Dunn, Michael John. "A Grammar of Chukchi." (1999).

It is *possible* to collect a large corpus of Chukchi transitive sentences, label a “subject”, “verb”, and “direct object” in each one, and look at which ordering turns out to be most common, but this tells you absolutely nothing useful about how Chukchi sentences are actually organized, for three reasons:

1. Transitive sentences with two noun phrases in them are, in general, extremely rare in Chukchi, and in no way “basic”.
2. Ordering of sentence elements is different depending on the discourse function of the sentence—i.e., does it have argument focus, predicate focus, or clause focus?
3. “Subject” may not actually be a useful grammatical category for describing Chukchi syntax anyway.

This highlights three assumptions underlying the Greenbergian list of six possible word order types, all of which can be broken:

1. That formally-transitive sentences (with overt arguments) are a significant part of a language’s grammar.
2. That basic transitive sentences will actually have at least three words in them (whether some of those words be nouns or pronouns).
3. That among those three words will be things you can easily identify as a “subject”, an “object”, and a “verb”.

And it turns out you only have to go as far as *German*—hardly an exotic language from an English-speaking perspective—to find an language which breaks a fourth assumption:

4. That the syntax of basic transitive sentences can be defined by a fixed relative ordering of those three elements

Recall that German syntax is most simply described as V2—i.e., the finite verb always appears second, regardless of what the subject, object, oblique arguments, or adverbial phrases are doing, and that other stuff can all move around. You could look at the statistical distribution of S, V, and O elements in German transitive sentences, and assign a most-common word order if you wanted to, much more easily than you could in Chukchi... but while it may be statistically true that the subject usually comes before the verb and the object usually comes after, that is far less useful than knowing the underlying rule that *independent finite verbs are always second*, and examining the implications which that has for where everything else goes, which are more extensive than just what happens with the subject and direct object.

Aside from being not particularly useful to do for German, it turns out that there are some languages where it is not possible to find a statistically-basic transitive word order at all—and not merely because subjects and object are difficult to identify, or

because sentences with more than one noun phrase (as opposed to, e.g., pronominal agreement affixes) are rare or restricted to a single discourse function, but because clauses with more than one noun phrase are strictly *ungrammatical!*² In that sort of situation, while it may be possible to determine a basic order for verbs vs. arguments, it is *impossible* to tell what the ordering of S vs. O would be, and so it is impossible to assign the language to one of the six basic word order buckets. And of course, this is all before we even consider languages in which it is difficult to tell the difference between a noun and a verb in the first place.³

And here's another wrinkle: even in languages like English, German, or French which allow multiple noun phrases in all sentence types (regardless of things like topic/focus marking, polarity, mood, voice, etc.), in natural speech pronouns are often much more frequent than full noun phrases anyway—and pronouns, light noun phrases, and heavy noun phrases can all trigger radically different word order behaviors! So if you claim that your language has, for example, a basic SOV word order, on the basis that a “basic” transitive sentence has two full noun phrases and a verb, you will fail to describe the actual ordering of elements in the sentences which actually occur most frequently! Even in English, there are tricky rules about how pronouns vs. full noun phrases can be ordered with respect to particles in phrasal verb constructions, but this behavior is extremely obvious in French: In theory, French is SVO, but in practice, direct and indirect object pronoun clitics cluster *in front* of the verb, and these clitic clusters are extremely common⁴. It still makes sense to describe French as *head-initial*, which implies a cluster of typological features alongside VO ordering, rather than labeling it SOV just because of the behavior of its pronouns, but if you started out designing a language specifically to be “SVO”, you could easily miss out on these kinds of complexities.

Now let's get back to looking at alignment. Supposedly, morphosyntactic alignment answers the question “how do you tell which thing is the subject and which thing is the object?”; or, more generally, “how do we know who is doing something to what?” But if you examine the five grouping options, only 3 of them—nominative-accusative, ergative-absolutive, and tripartite—actually manage to provide unambiguous answers to that question, just based on the formal marking groups. And yet, other alignments (notably direct alignment)⁵ do actually exist in natural languages, which should be your first clue that something else is going on—there must be more to alignment than just how you mark the arguments of a clause, and that opens up the

² You might reasonably wonder how you could express a sentence like “Jim ate spaghetti.” in this kind of language. In most discourse situations, you simply *wouldn't*—either Jim or the spaghetti, or both, would've been brought up previously, such that you could replace one or both of them with a pronoun. But if you did need to explicitly reference both Jim and his spaghetti, you would need a biclausal construction that chooses one of those referents to activate as the discourse topic: “This is Jim; he ate spaghetti”, “Spaghetti was eaten; Jim did it.”, “Jim ate it; it was spaghetti.”, etc.

³ Kearsley, Logan. “How To Not Verb”. *Fiat Lingua*, Apr. 2020

⁴ Some people would argue that in modern French, these aren't actually “pronouns” at all, but polypersonal agreement prefixes—but even so, differences in word order between pronouns and full noun phrases show up in lots of languages to varying degrees. See, for example, Spanish “*Lo amo.*” vs. “*Amo al gato.*”

⁵ Even the transitive or Monster Raving Loony alignment actually does occur in Rushani, an Iranian language spoken in Afghanistan and Tajikistan, as part of a split alignment system in the past tense.

door to a lot more options. For example, direct-inverse (also known as “hierarchical alignment”) systems make use of a lexical hierarchy⁶ to determine what the subject and object are based on the properties of the words used to refer to them in any given sentence, with marking on the *verb* to indicate whether the assignment of subject and object should conform to normal expectations, or be inverted. Other direct and transitive alignment languages may rely on discourse-specific pragmatics and/or multi-clausal constructions to clarify argument roles, or rely on associating subjects with *topics* and objects with *focuses* (a grammatical restriction that can also be paired up with other alignment systems, and drive the development of grammatical voice transformations; English, for example, does not require subjects to be topics, but does prefer that, which is a common reason to use passive voice when the topic would otherwise be a morphosyntactic object). There probably aren’t any natural languages that do *only* this (although Thomas Payne cites Sierra Popoluca as a possible example in *Describing Morphosyntax*), but it comes up frequently as part of a mixed strategy in languages that otherwise have “normal”-but-deficient alignment marking; e.g., many nominative-accusative languages neutralize the nominative and accusative cases for some declension classes (usually corresponding to neuter gender or inanimate referents), and languages with hierarchical alignment sometimes run into sentences with two argument in the same position on the hierarchy—and in the absence of some other mixed-in system, topicality disambiguates.

It is also very common to mix alignment systems between different parts of a single language. “Split ergativity” is a blanket term for a variety of different systems which sometimes use ergative marking and sometimes use accusative marking, depending on the type of sentence you have, where the difference could be triggered by tense, aspect, or even whether or not there is a third-person pronoun in the clause! Active-stative or semantic alignment, also called fluid-S alignment, alters the marking class of the subject between ergative-absolutive and nominative-accusative based on volitionality—and this kind of pattern is sometimes seen mixed into otherwise obviously nominative-accusative languages in the form of “quirky case” or “impersonal constructions”, like in Icelandic or Russian. Also keep in mind that alignment is about how marking types are grouped, but is independent of how that marking is done—fixed word order and thorough morphological case marking are both “marking” for purposes of identifying morphosyntactic alignment, and this is another place where alignment can be split. Many “ergative” languages have ergative *morphological* marking, but nominative *syntax*—e.g., SVO order for transitives, and SV for intransitives, rather than the ergative VS pattern, or cross-clause argument sharing that identifies common arguments based on shared subject status rather than shared case—which of course requires that “subject” be a meaningful category for syntax purposes, even if it isn’t for case marking.

If you *can’t* reliably identify a “subject” as a meaningful syntactic category, then making any kind of distinction based on how “subjects” are marked or shared between sentences becomes meaningless, and we find that we have moved beyond the

⁶ Usually an “animacy hierarchy”.

five-category system entirely! Austronesian alignment provides an example of a more complex type of alignment-mixing; the simplest form of Austronesian alignment can be seen as a mixture of nominative-accusative and ergative-absolutive—or in other words, a split-ergative system—with a freely-chosen inflection on the verb determining which marking system you use in any given sentence. More complex Austronesian or symmetric-voice systems end up simply not fitting into the S-A-O grouping categories at all, despite having identifiable subject—or, perhaps, fitting nicely into one of them from the perspective of formal argument marking, but with a whole extra system of verbal inflection that you need to pay attention to in order to figure out the actual *semantics* of those argument markers in any given sentence.

How might we as conlangers explicitly break out of these categorizations on purpose, aside from copying languages that already do so? Well, both the Greenbergian basic word order system and the 5-bucket morphosyntactic alignment system rely on the concept of a “subject”, so an easy way to break both would be to avoid having a “subject” in your syntax at all. Previously, I mentioned that Chukchi does exactly that—so how does Chukchi manage it? Chukchi grammar cares a lot about a morphosyntactic slot which Dunn’s grammar labels as “the absolutive”. It’s the only required argument for most sentences, even if there is also an “ergative” participant, but it’s also the only slot where you can place a complete noun phrase (other non-pronominal arguments can be at most a single unmodified noun, without any adjectives or relative clauses attached), and it’s used for focused, not topical, material.

More generally, we can ask “what actually *is* a subject”? And then just do... not that. Defining subjects in a cross-linguistically useful way turns out to be non-trivial, but we can identify most *prototypical* subjects as follows:

1. Triggers agreement on a verb.
2. In basic sentences (i.e., without explicit voice/valency changes), typically refers to semantic agents or themes.
3. Typically aligns with the discourse topic.
4. Exists outside of the verb phrase in syntax.
5. If the language allows omitting any arguments, the subject will be omissible.
6. If the language *requires* at least one argument, it will be the subject.
7. If coordinated clauses can share an argument, that argument is the subject.

Not all languages will have subjects that meet all of these criteria (even though most natural languages *do* have subjects), but the fewer such criteria that are met by any single category in the grammar, the less subject-like it is, until it no longer makes sense to talk of a “subject” at all because the functions are too much divided among different morphosyntactic categories. Chukchi manages to chuck out more than half of all of those requirements, so “subject” is not a useful category for describing Chukchi!

How could we eliminate subjects without just copying Chukchi? Suppose you have completely flat syntax and no agreement and polypersonal agreement, so that no

argument can be said to uniquely control the verb or be distinguished from the other by being outside or inside the verb phrase. You're well on your way to having a language with no subject!

Suppose you don't have verbs—or, more strongly, *predicate phrases*—at all. Well, now you've broken *another* part of the basic word order system, but you also can't have a subject of a verb if there is no verb!⁷

Suppose you use strict semantic marking of all arguments, rather than purely-formal cases, such that there is no “nominative”, “accusative”, “ergative”, or anything like that—just unique and invariant marking for “agent”, “patient”, “theme”, “goal”, “experiencer”, “location”, etc. At first glance, this might look like some mix of fluid-S and tripartite marking, but the markings will be potentially different for every sentence—that's another way to break out of both the word order and alignment boxes. Or perhaps the marking is not completely semantically determined, but lexically specified for every single verb—either every verb has its own alignment, or that's just a totally new alignment system not covered by current typology. Either way, you can't squish the whole language into a pre-defined box!

However, my main goal here is not to tell you that you *must* tear down typological barriers and do weird things. Take some of those suggestions if you like, and there are plenty of other approaches possible, but I shall avoid giving further examples here, because the real goal is just to show that you don't need to worry about the boxes. They were always descriptive, not prescriptive, and they were always imprecise and insufficient. To create more varied and interesting languages, you don't have to look at the boxes and think of how to break out of them—just ignore the boxes completely! Stop thinking of terms of “I will design a language with X word order and Y alignment”, and start thinking in terms of “this is how I will indicate who is doing what in any given sentence type, and what it means, if anything, to put the elements of a clause in a particular order.” You, or anyone else, can do the analysis to put the resulting system into basic word order and morphosyntactic alignment boxes later, and if it turns out not to fit neatly into any of them... that's fine! If the system works, it does not matter if you can't squeeze it into a typological box.

⁷ See again: “How To Not Verb”. *Fiat Lingua*, Apr. 2020