

# Verbal Predication Constructions in Universal Dependencies

**William Croft**

University of New Mexico  
Department of Linguistics  
wacroft@icloud.com

**Joakim Nivre**

Uppsala University  
Department of Linguistics and Philology  
joakim.nivre@lingfil.uu.se

## Abstract

Is the framework of Universal Dependencies (UD) compatible with findings from linguistic typology about constructions in the world’s languages? To address this question, we need to systematically review how UD represents these constructions, and how it handles the range of morphosyntactic variation attested across languages. In this paper, we present the results of such a review focusing on verbal predication constructions. We find that, although UD can represent all major constructions in this area, the guidelines are not completely coherent with respect to the criteria for core argument relations and not completely systematic in the definition of subtypes for nonbasic voice constructions. To improve the overall coherence of the guidelines, we propose a number of revisions for future versions of UD.

## 1 Introduction

Universal Dependencies (UD) is a framework for morphosyntactic annotation, which is designed to be applicable to all human languages in a way that enables meaningful cross-linguistic comparisons (Nivre et al., 2016, 2020; de Marneffe et al., 2021). Construction grammar has also been combined with linguistic typology to allow for cross-linguistic comparison of grammatical constructions (Croft, 2016, 2022). This paper contributes to the project of adding the third edge to this triangle: representing cross-linguistically valid constructions in UD (Nivre, 2025). To find out whether UD can represent typologically justifiable constructions, Nivre (2025) proposes to build a construction for UD based on the survey of universal constructions and morphosyntactic realization strategies in Croft (2022) and the MoCCA database of comparative concepts derived from it (Lorenzi et al., 2024).

Croft’s survey is based on two types of comparative concepts (Haspelmath, 2010; Croft, 2016):

*constructions*, which are universal form-function pairings defined solely in terms of their function, and *strategies*, which are non-universal and defined by the pairing of a function with some cross-linguistically identifiable morphosyntactic form. Annotations in UD are not defined in terms of constructions and strategies, but for the framework to be universally applicable it must be possible to annotate all major constructions and strategies in the world’s languages. And to support cross-linguistic comparisons, these annotations should ideally reflect systematic correspondences in constructions and strategies across languages.

The research program outlined in Nivre (2025) is to develop a construction for UD, consisting of the following components:

- An inventory of universal constructions.
- For each construction, an inventory of common strategies for realizing that construction in the world’s languages.
- For each construction-strategy pair, a cross-linguistically valid UD analysis and representative examples from different languages.

This will help improve cross-linguistic annotation consistency by providing a complementary view of the UD guidelines, which is holistic and onomasiological; it will also provide better support for construction-based annotation on top of UD (Weissweiler et al., 2024); it will finally reveal to what extent UD can represent constructions and strategies systematically and transparently across languages, thereby identifying shortcomings in the current guidelines.

The first contribution to this project can be found in Nivre and Croft (2025) and reviews the guidelines for reference and modification constructions in UD. In this paper, we proceed to discuss verbal predication constructions, or verbal clauses, involv-

ing simple verbal predicates and their arguments. This family of constructions is discussed in Chapters 6–9 of Croft (2022).

## 2 Verbal Clause Constructions

A verbal clause construction consists of two types of elements: the head, which is a verb denoting an action or event, and argument phrases denoting participants of the action or event. This is exemplified in (1), from Croft (2022, p. 180), where the verb *broke* is associated with four argument phrases: *Sue*, *a coconut*, *for Greg*, and *with a hammer*.

- (1) Sue broke a coconut for Greg with a hammer.

The grammatical encoding of argument phrases is primarily determined by their degree of salience or *topicality* to the interlocutors in the discourse. The most topical argument is encoded by the *subject*, the next most salient argument by the *object*, and all other arguments by *oblique* phrases. For example, in (1), *Susan* is the subject, *a coconut* is the object, and *for Greg* and *with a hammer* are obliques. Subjects and objects are often grouped together as *core arguments*.

In the most prototypical clause constructions, the more topical arguments are also the more central participants of the action or event. Thus, in (1), the subject denotes the agent of the action, and the object denotes the object most directly affected by the action, while the oblique arguments denote more peripheral participants. Such constructions are called *basic voice* constructions and are discussed in Section 3. In Section 4, we then turn to constructions that have been conventionalized to express non-prototypical combinations of participant roles and argument salience.

## 3 Basic Voice Constructions

Basic voice constructions are traditionally classified based on the number of central participant roles, or core arguments, into intransitive, transitive, and ditransitive constructions. We will begin with the transitive construction, with two core arguments, which is generally assumed to be the prototypical verbal clause (Croft, 2022, p. 183).

### 3.1 The Transitive Construction

If the transitive construction is the most prototypical verbal clause construction, the most prototypical event type expressed through this construction is an agentive change of state event, that is, an event

where an external volitional agent brings about a change in a patient. The asymmetric semantic relation between agent and patient is force-dynamic, that is, the change of state event involves a transmission of force from the agent to the patient (Talmy, 1988; Croft, 2010). To facilitate cross-linguistic comparison, typologists have proposed that the construction be defined by an even more specific event type, the agentive breaking event exemplified in (1) (Haspelmath, 2011, 2015; Croft, 2022).

In the prototypical transitive clause, the phrase expressing the agent (A) role is the subject, and the phrase expressing the patient (P) role the object. But the same construction is commonly used also to express other event types with other semantic roles, such as motion events or experiential events. Thus, in a sentence like *she entered the cave*, the subject (*she*) expresses the figure role (F), and the object (*the cave*) expresses the ground role (G). And in *she saw the sun*, the subject (*she*) is an experiencer (X) and the object (*the sun*) is a stimulus (M).

There are cross-linguistic generalizations about the tendency for different event types to recruit<sup>1</sup> the transitive construction, often summarized in so-called transitivity hierarchies (Tsunoda, 1981, 1985; Malchukov, 2005; Beavers, 2011). To map out the distribution of the transitive construction in a given language, we need to study how the subject and object are encoded in prototypical transitive clauses and see to what extent the same encoding appears with other event types and semantic roles. Generally speaking, there are three common strategies used to distinguish arguments in verbal clauses, including transitive clauses, exemplified in (2–4) (Croft, 2022, pp. 187–188).

- (2) Tanj-a                      ubi-la                      Mašu  
Tanya-F.NOM kill-PST:FSG Masha-F.ACC  
'Tanya killed Masha'
- (3) x-Ø-uu-choy                                      chee7 tza7n ikaj  
PST-3SG.ABS-3SG.ERG-cut tree    with axe  
'he cut tree(s) with an axe'
- (4) ka'se'kaw: samlap ko:n kru:k  
farmer    kill    child pig  
'(the) farmer(s) kills/killed (the) piglet(s)'

The Russian example (2), from Comrie (1989), exemplifies the use of *flags*, morphemes that encode the semantic relationship between the participant

<sup>1</sup>Recruitment is a relationship between two constructions in which the structure of one construction is recruited for use, or extended to use, in the other construction (Croft, in press).

and the event. In this example, flags take the form of case affixes, but flags can also be realized as adpositions. Cross-linguistically, there is a tendency for argument phrases lacking overt flags to express core argument roles.

The Tzutujil example (3), from Dayley (1989), illustrates the strategy of *indexation*, where an argument is indexed by a morpheme that is typically an affix of the predicate. In this case, both the subject and the object are indexed on the verb. Cross-linguistically, there is a strong tendency that indexed arguments express core argument roles.

Since flags occur on arguments and indexation occurs on the predicate, the two strategies may be used together. This is the case in the Russian example (2), where the subject *Tanja* carries a flag and is also indexed on the verb.

The Khmer example (4), from Haiman (2011), uses neither flags nor indexation, and the arguments are distinguished only through *word order*. The cross-linguistic study of basic word order in transitive clauses goes back to Greenberg (1966) and has shown that it is overwhelmingly more common for subjects to precede objects in languages that have a dominant order.

### 3.2 The Intransitive Construction

The intransitive clause construction involves a verb and a single core argument, whose role is called S by typologists, which is almost always encoded like one of the two core arguments in the transitive construction. The encoding patterns of the three roles is called *alignment* and the three most common patterns are *neutral* ( $A = S = P$ ), *accusative* ( $A = S \neq P$ ) and *ergative* ( $A \neq S = P$ ) alignment. Neutral alignment is found in English when no argument is realized as a pronoun, as shown in example (5). Example (6), from Weber (1989), shows accusative alignment in Huallaga Quechua, involving both flags and indexation; example (7), from Williams (1980), shows ergative alignment with flags in Yuwaalaraay.

- (5) a. the dog barked  
b. the dog chased the cat
- (6) a. yaku-Ø timpu-yka-n  
water.NOM boil.IPFV-3  
'the water is boiling'  
b. Hwan-Ø Tumas-ta maka-n  
John.NOM Tom.ACC hit.-3  
'John hits Tom'

- (7) a. wa:l nama yinar-Ø banaga-ni  
NEG that woman-ABS run-NFUT  
'the woman didn't run'  
b. duyugu nama dayn-Ø yi:-y  
snake-ERG that man-ABS bite-NFUT  
'the snake bit the man'

Regardless of the alignment, however, the single core argument in an intransitive clause is classified as a subject, because it is the single most topical argument of the construction.

### 3.3 Reflexives and Reciprocals

In addition to the transitive and intransitive constructions, there are two constructions that have affinities with both and often employ the same strategies: the *reflexive* and the *reciprocal* construction. The reflexive construction expresses an event with a single participant (like intransitives) but two distinct roles (like transitives), as in *she injured herself*. The reciprocal construction expresses an event with a pair of participants that both assume the same two roles, as in *they touched each other*.

Both reflexives and reciprocals typically recruit either the transitive or the intransitive construction for their realization. The former is the *dual-role* strategy, with two argument phrases, as in the examples above, and may involve a specialized argument expression such as a reflexive or reciprocal pronoun. The latter is the *single-role* strategy, with only one argument phrase, as in *he shaved* and *they met*. Cross-linguistically, the dual-role strategy often grammaticalizes into the single-role strategy through fusion of a specialized reflexive/reciprocal element with the verb (Croft, 2022, pp. 208–209).

### 3.4 The Ditransitive Construction

The ditransitive clause construction is defined in terms of *transfer events*, physical transfer events expressed by verbs like *give* and *sell*, as well as mental transfer events expressed by verbs like *show* and *tell*. The roles associated with these events are agent (A), theme (T), and recipient (R). There is a force-dynamic ordering  $A > T > R$ , but R is almost always human and hence topical enough to be encoded as a core argument, comparable to T. While the A role appears to be universally encoded as the grammatical subject, languages use different strategies for encoding the T and R roles, which can again be described in terms of alignment with the transitive construction. In the *neutral* alignment ( $T = P = R$ ), or double-object strategy, both

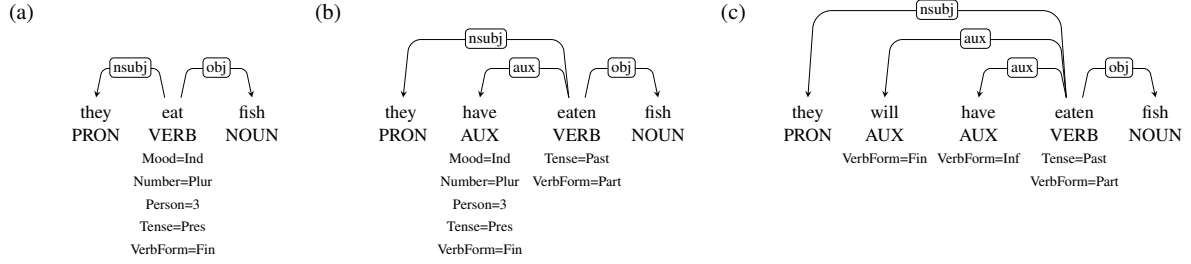


Figure 1: UD annotation of verbal predicates: (a) finite main verb, (b–c) main verb with auxiliaries.

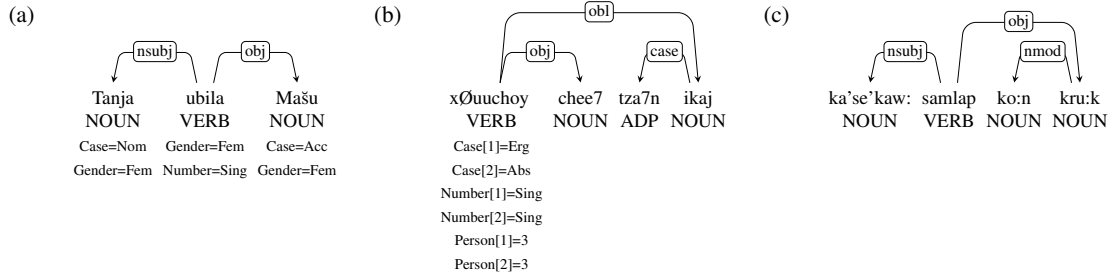


Figure 2: UD annotation of encoding strategies: (a) flags, (b) indexation, (c) word order.

T and R are co-expressed with P. In the *indirective* alignment ( $T = P \neq R$ ), T and P are co-expressed and referred to as the *direct* object, while R has a distinct encoding and is referred to as the *indirect* object. In the *secundative* alignment ( $T \neq P = R$ ), finally, R and P are co-expressed and distinct from T. In this case, the phrase expressing R or P is the *primary* object, while the phrase expressing T is the *secondary* object.

### 3.5 UD Annotation

When reviewing the UD annotation of basic voice constructions, our discussion will focus on how UD treats different *alignment* strategies across intransitives, transitives and ditransitives. Before we turn to that discussion, however, we will briefly review how UD annotates *verbal predicates* and how it handles the *encoding strategies* used to distinguish arguments in any of these constructions.

#### Verbal Predicates

The predicate of a verbal clause consists of a main verb, which is assigned the part-of-speech tag VERB, possibly together with one or more auxiliaries, which are assigned the tag AUX. Both main verbs and auxiliaries may be assigned morphological features capturing properties such as tense, mood, and aspect. It is worth noting that UD always treats the main (lexical) verb as the root of the clausal structure, regardless of whether it is

finite or not, and attaches auxiliaries to the main verb with the syntactic relation *aux*, as illustrated in Figure 1.<sup>2</sup>

#### Encoding Strategies

As observed in Section 3.1, there are three main strategies used to distinguish arguments: flags, indexation, and word order. These are annotated to varying degrees in UD, using part-of-speech tags, morphological features, and relations:

- Flags realized as morphological affixes are represented by the morphological feature Case, as shown for example (2) in Figure 2(a), while adpositions are tagged ADP and attached with the *case* relation, as exemplified by the oblique argument in Figure 2(b).
- Indexation is also represented by morphological features, whose values correspond to those of the indexed arguments, as shown for example (3) in Figure 2(b). When multiple arguments are indexed, as in this example, the technique known as *layering* is used to represent multiple values of the same feature. However, as observed by Nivre and Croft (2025), there

<sup>2</sup>In these and all following examples, we simplify the UD representations by omitting (a) lemmas and (b) morphological features that are not relevant for discussion (notably features on nominal arguments in these examples and features on verbal predicates in subsequent examples).



is nothing in the annotation that explicitly connects the index features to the arguments.

- The word order strategy is not annotated explicitly, but word order is preserved in the representation; cf. example (4) and Figure 2(c).

### Intransitive–Transitive Alignment

The intransitive construction is annotated in UD by attaching the single core argument to the verb with the *nsubj* relation. This is consistent with the analysis in Croft (2022) in that the phrase expressing the S role is analyzed as the grammatical subject.

For the transitive construction, the idea is to use *nsubj* and *obj* for any two arguments encoded as the A and P arguments of a prototypical transitive clause describing an agentive change-of-state event, including clauses describing motion events (*she entered the cave*) and experiential events (*she sees the sun*). However, if one of the arguments has an oblique encoding, then it is instead annotated with the *obl* relation, even if it expresses the same role as in the corresponding transitive clause. Thus, a clause like *she ran into the cave* is analyzed as an intransitive clause, with *she* as *nsubj* and *into the cave* as *obl*, and similarly for a clause like *she looked at the sun*. Of course, oblique arguments may also appear in transitive clauses, as in caused motion events like *she chased them into the cave* (*she* = *nsubj*, *them* = *obj*, *into the cave* = *obl*).

The question, however, is how to identify subjects and objects in languages with different alignment strategies. The documentation on the UD website<sup>3</sup> appears to follow Croft (2022) in treating the phrase expressing the A role in a prototypical transitive clause as the grammatical subject regardless of alignment, because it is the most topical argument. More specifically, it says that “case alignment should not be used to decide the assignment of core argument roles” and that “in ergative languages, the patient-like argument of a transitive verb (O/P) will take the *obj* relation despite the fact that it carries the same case marking as the *nsubj* argument (S) of an intransitive verb”. The annotations in Figure 2 are compatible with these guidelines, specifically Figure 2(b), where the argument indexed with absolutive case is analyzed as *obj*. However, in a more detailed discussion of ergativity, de Marneffe et al. (2021) argue that, while this analysis is appropriate for languages where

ergative–absolutive case marking is primarily a morphological feature, such as Basque, there are other languages, such as Jirrbal (or Dyirbal), where ergativity extends to syntactic relations. For such languages, de Marneffe et al. (2021) propose an analysis based on Dixon (1994), where the S and P arguments are treated as a “pivot” and are both assigned the *nsubj* relation, while the A argument is instead assigned the *obj* relation. To indicate the unusual role assignment, it is recommended to use the subtype *nsubj:pass*<sup>4</sup> for the P argument and the subtype *obj:agent* for the A argument (de Marneffe et al., 2021, p. 295).

### Reflexives and Reciprocals

Reflexive and reciprocal constructions are in principle annotated exactly as the constructions they recruit, that is, the transitive or intransitive construction. However, if a language employs a specialized dual-role strategy involving a reflexive or reciprocal pronoun, this may be captured by features on the pronoun, such as Reflexive=Yes and PronType=Rcp. The UD guidelines also prescribe a special treatment of so-called inherent reflexive verbs, such as *se souvenir* (remember) in French, where the verb cannot occur with a non-reflexive pronoun and where there is arguably only one semantic role. In this case, the reflexive pronoun should be attached to the verb with the *expl* (expletive) relation (instead of the *obj* relation) to indicate that it does not express a semantic role in relation to the predicate.

### Ditransitive–Transitive Alignment

UD defines ditransitive clauses more narrowly than Croft (2022) and only has specific guidelines for the neutral alignment strategy, where the T and R roles are both encoded as core arguments. In this case, UD assigns *nsubj* to the A argument, *obj* to the T argument, and a special relation *iobj* (for indirect object) to the R argument. The *obj/iobj* distinction is upheld even if the T and R arguments have identical encoding, and is thus based on roles rather than morphosyntactic realization.

For the *indirective* strategy, UD uses *nsubj* for the A argument, *obj* for the T argument, and *obl* for the R argument with an oblique encoding, which typically involves either an adposition, as in English *she gave the book to Peter*, or morphological

<sup>3</sup><https://universaldependencies.org/u/overview/simple-syntax.html#intransitive-and-transitive-clauses>

<sup>4</sup>The subtype *:pass* was first used in the analysis of passive constructions (hence the name), but it is now used more generally in UD for subjects whose semantic role is lower than expected in the transitivity hierarchy.

Construction	Strategy	Roles					
		S	A	P	T	R	C
Intransitive	–	nsubj					
Transitive	Accusative		nsubj	obj			
	Ergative 1		nsubj	obj			
	Ergative 2		obj:agent	nsubj:pass			
Ditransitive	Neutral		nsubj		obj	iobj	
	Indirective		nsubj		obj	obl	
	Secundative		nsubj		obl	obj	
<b>Construction</b>	<b>Basic Voice</b>						
Passive	Transitive		obl:agent	nsubj:pass			
Causative	Intransitive		obj:caus				nsubj
	Transitive		iobj:caus	obj			nsubj

Table 1: UD relations for semantic roles in verbal clause constructions (C = external causer).

case, as in Latin *librum Petro dedit* (he/she gave the book to Petrus), where the oblique R argument *Petro* is in dative case, while the object *librum* is in accusative case.<sup>5</sup> The *secundative* strategy is not described in the UD guidelines, but it is natural to assume that the core R argument is annotated *obj* (since the *iobj* relation normally requires the presence of an *obj* argument in the same clause), while the oblique T argument is annotated *obl*.

### Interim Summary

The upper part of Table 1 summarizes the UD treatment of basic voice constructions by showing how prototypical semantic roles are mapped to syntactic relations (with the two different treatments proposed for transitives with ergative alignment).

## 4 Non-Basic Voice Constructions

Non-basic voice constructions are clausal constructions used to express a non-prototypical combination of the topicality of referents and the participant roles those referents play in the event denoted by the predicate.

### 4.1 Passive–Inverse Constructions

A passive–inverse construction expresses a situation where the P referent has higher topicality than the A referent (Croft, 2022, p. 252). In the English passive construction in (8b), the P argument (*he*) is coded like the A argument (*she*) in the prototypical active construction in (8b), while the A argument

(*by her*) is oblique. In the Algonquian inverse construction in (9b), from Wolfart and Carroll (1981), the P argument is again coded like the A argument in the direct construction in (9a), but the A argument is now coded as the P argument in the more prototypical construction.

- (8) a. she took him to school  
b. he was taken to school (by her)
- (9) a. ni-wāpam-ā-wak  
1-see-DIR-3PL  
‘I see them’  
b. ni-wāpam-ikw-wak  
1-see-INV-3PL  
‘they see me’

These are only two of the many strategies used in passive–inverse constructions in the world’s languages. For further discussion, see (Croft, 2022, pp. 256–263).

### 4.2 Antipassive Constructions

Antipassive constructions involve a P argument with lower topicality than in a basic transitive clause. Such constructions are common in ergative languages, where the P argument is demoted to an oblique and the A argument takes over the absolutive encoding. Example (10), from Patz (2002), illustrates the antipassive construction in Kuku Yalanji.

- (10) a. nyulu dingkar-angka minya-Ø nuka-ny  
3SG.NOM man-ERG meat-ABS eat-PST  
‘the man ate meat’  
b. nyulu dingkar-Ø minya-nga muka-ji-ny  
3SG.NOM man-ABS meat-LOC eat-ANTP-PST  
‘the man had a good feed of meat’

<sup>5</sup>A dative case argument may be treated as a core argument, hence *iobj*, if other criteria point to it being core, notably if it is indexed on the verb.

Example (10) illustrates the oblique P strategy, which is also found in an English example like *the dog chewed the bone* versus *the dog chewed on the bone* (although without overt coding on the verb). Other common strategies in antipassive constructions are the omitted P strategy (*she ate a sandwich* versus *she ate*) and different types of noun incorporation (Croft, 2022, pp. 266–270).

### 4.3 Causative Constructions

Causative constructions add an external causer (C), universally encoded as a subject core argument. The encoding of the ordinary subject, the causee, depends on what strategy is used, and sometimes also on whether the base clause is transitive or intransitive. Many languages use a *complex predicate* strategy, as in the English examples in (11), where the causee becomes the direct object and everything else stays the same.

- (11) a. she made him cry  
b. she made him write the letter

Turkish instead uses a *simple predicate* strategy, with overt coding on the verb, as shown in (12), from Comrie (1989). In (12a), the base clause is intransitive and the causee is expressed as an object with accusative encoding; in (12b), the base clause is transitive and the causee is expressed as an oblique with a dative flag.

- (12) a. Ali Hasan-ı öl-dür-dü  
Ali Hasan-ACC die-CAU-PST  
'Ali killed Hasan'  
b. Dişçi mektub-u müdür-e imzala-t-tı  
dentist letter-ACC director-DAT sign-CAU-PST  
'the dentist made the director sign the letter'

### 4.4 Applicative Constructions

In applicative constructions, a peripheral participant is encoded as a core argument, usually as an object, and the object of the corresponding prototypical transitive clause may be encoded as an oblique. This is illustrated with a Hungarian example in (13), from Moravcsik (1978).

- (13) a. János fák-at ültetett a kert-be  
John trees-ACC planted the garden-into  
'John planted trees in the garden'  
b. János be-ültette a kerte-t fák-kal  
John APPL-planted the garden-ACC trees-with  
'John planted the garden with trees'

Hungarian uses a simple predicate strategy, with overt coding on the verb, but it is also common to use a complex predicate strategy for applicative constructions, in particular a serial verb strategy.

## 4.5 UD Annotation

### Passive–Inverse Constructions

Passive constructions are annotated in UD by attaching the passive subject to the verb with the subtype relation *nsubj:pass* to indicate that it expresses the argument role associated with the direct object in the corresponding transitive clause. The agent phrase, if present, is annotated using the subtype *obl:agent*. If the verb is overtly marked for the passive voice, it carries the feature *Voice=Pass*; if the passive is a periphrastic construction, the auxiliary may instead be annotated with the subtype *aux:pass*. Inverse constructions like the one in (9b) are not discussed explicitly in the UD guidelines, but it seems straightforward to use the subtypes *nsubj:pass* and *obj:agent* recommended for transitive clauses in (some) languages with ergative alignment, with the feature *Voice=Inv* on the verb.

### Antipassives

Antipassives are not explicitly discussed in the UD guidelines, but the oblique P and omitted P strategies can be straightforwardly annotated using the existing guidelines.<sup>6</sup> The treatment of noun incorporation in UD is a more controversial issue, which we will sidestep in this paper. We refer the interested reader to Tyers and Mishchenkova (2020) for a discussion and a proposal.

### Causatives

For causatives with the simple predicate strategy, UD recommends using the subtypes *obj:caus* and *iobj:caus* for the causee, as shown in Figure 3 for the Turkish examples in (12) (with the feature *Voice=Cau* on the verb). The use of the *iobj* relation here is unexpected, given that the argument has an oblique encoding, and the subtype *obl:caus* would seem more natural. For the complex predicate strategy, illustrated by the English example (11), the causee will normally be annotated with the *obj* role (without subtype), while the second verb will be assigned the *xcomp* relation.

<sup>6</sup>In the former case, the demoted P argument is assigned the *obl* relation; in the latter case, it is simply dropped.

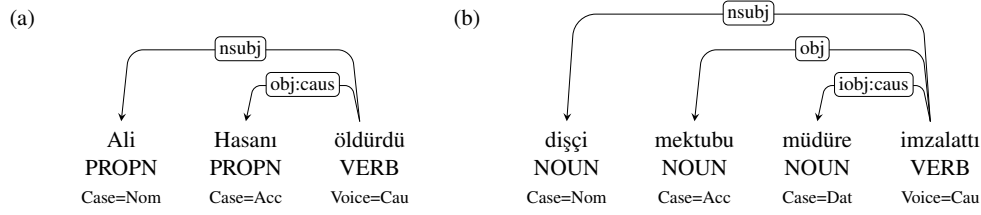


Figure 3: UD annotation of causative constructions.

## Applicatives

Applicatives do not appear in the official UD guidelines, but there is a short discussion in [de Marneffe et al. \(2021\)](#) of ditransitive applicatives in Swahili, where it is recommended to use the *iobj* relation for the promoted argument if it is indexed by the verb. For the Hungarian example (13b) it seems natural to use *obj* for the promoted argument and *obl* for the demoted one. In addition, one could envisage a feature *Voice=App* on the verb, but no such feature currently exists in UD.

## Interim Summary

The treatment of nonbasic voice constructions in UD is summarized in the lower part of Table 1. We have, however, only included constructions for which there are official guidelines.

## 5 Discussion

Our review has shown that the UD annotation framework can in principle represent all the major constructions and strategies for verbal predication discussed in [Croft \(2022\)](#), even though not all non-basic voice constructions are treated explicitly in the current documentation of the UD guidelines. These guidelines are summarized in Table 1, which can be regarded as a blueprint for the UD construction of verbal predication constructions.

However, we have also observed a few cases where the UD treatment does not quite align with comparative concepts from typology, and sometimes arguably even conflicts with basic principles of UD itself. One such case is the treatment of transitives in ergative languages, where [de Marneffe et al. \(2021\)](#) advocates a mixed analysis, which is sometimes based on topicality, sometimes on morphosyntactic encoding, specifically case alignment. Another case is the analysis of ditransitive clauses with neutral alignment, where the use of the *iobj* relation appears to be motivated on different grounds than other core argument relations. Finally, we note that the use of subtypes to mark

non-prototypical argument realizations can be improved with respect to systematicity and naming conventions. Nevertheless, we believe that, with relatively small adjustments, the guidelines can be made globally coherent and consistent with basic UD principles as well as findings from linguistic typology. We will now try to outline these modified guidelines and their motivation.

A cornerstone of UD is the assumption that the core-oblique distinction, albeit not completely unproblematic, is a better foundation for morphosyntactic annotation than the argument adjunct distinction ([de Marneffe et al., 2021](#), pp. 266–268). The basis for distinguishing core arguments in a given language is the encoding of the two arguments in a prototypical transitive clause; any argument that uses the same encoding as one of these is core; any argument that uses a different encoding is oblique.

The basis for assigning specific syntactic relations to core arguments in *basic* voice constructions is topicality, with the *nsubj* relation reserved for the most topical argument and the *obj* relation for the second most topical argument. It follows that the single S argument in intransitive clauses is *nsubj*.

In transitive clauses, we assume that the topicality hierarchy is  $A > P$ , which means that the A argument is *nsubj* and the P argument *obj*, regardless of case marking or other coding properties, and all other arguments are *obl*. This analysis naturally carries over to other event types like motion events:  $F = nsubj$ ,  $G = obj$  in uncaused motion (*she entered the cave*);  $A = nsubj$ ,  $F = obj$ ,  $G = obl$  in caused motion (*she chased them into the cave*). For experiential events, the analysis mirrors the encoding in the prototypical transitives, which means that  $M = nsubj$  and  $X = obj$  in the causative construal (*she frightens them*) and vice versa in the attending construal (*they fear her*).

In ditransitive clauses, we assume that topicality reflects the force-dynamics ( $A > T > R$ ), which means that A is *nsubj* and that T is *obj* if it is realized as a core argument; the expected realization of



R is *obl*, which makes the indirective alignment the basic voice construction for ditransitives. We will therefore treat the neutral and secundative alignments as nonbasic voice constructions (more precisely as applicative constructions), which obviates the need for the *iobj* relation.

In *nonbasic* voice constructions, which by definition involve some kind of mismatch between topicality and encoding, we use subtypes to indicate deviances from prototypical argument realizations. Here we propose a new subtyping system based on the argument roles used in linguistic typology, including at least *:s*, *:a*, *:p*, *:t*, *:r*, and *:c* (*c* for causer). We believe that this will be a more expressive and coherent subtyping system than the current use of *:pass*, *:agent*, and *:caus*, which mixes different naming conventions (constructions vs. roles) and where especially *:pass* has a misleading name as it covers more than just passives. Given these subtypes, we can annotate nonbasic voice constructions transparently as follows:

- **Passive–Inverse:** The P/T/R argument is *nsubj:p/nsubj:t/nsubj:r*, and the A argument is *obj:a* or *obl:a*, depending on strategy.
- **Antipassive:** The A argument is *nsubj* and the P argument is *obl:p*.
- **Causative:** The causer is *nsubj:c*. If the base clause is intransitive, the S argument is *obj:s*; if the base clause is transitive, then the A argument is *obj:a* or *obl:a* and the P argument *obj* or *obl:p*, depending on strategy.
- **Applicative:** The A argument is *nsubj*, the P/T argument is *obj* or *obl:p/obl:t*, depending on strategy, and the promoted argument is *obj* with a subtype reflecting its role. A special case of this is a ditransitive with neutral or secundative alignment, where the R argument is *obj:r* (instead of *iobj*) and the T argument is *obj* (neutral) or *obl:t* (secundative).

A possible alternative to using role-based subtypes is to use a simpler system with only two general subtypes, *:high* and *:low*, which indicate that an argument has, respectively, higher or lower topicality than expected. The P argument would then be *nsubj:low* in a passive–inverse construction and *obl:high* in an antipassive construction. However, this would be a much less expressive system, which would make some nonbasic voice constructions indistinguishable (for example, inverse constructions and intransitive causatives).

Finally, and regardless of whether future versions of UD will adopt our proposed revisions of the annotation guidelines, there will be a need for additional morphological features to capture nonbasic voice constructions coded on the verb itself. This includes at least a feature or feature value for applicative constructions.

## 6 Conclusion

In this paper, we have taken another step towards a construction for UD, in the sense of Nivre (2025), by reviewing the way UD annotates constructions and strategies for verbal predication, following the taxonomy of Croft (2022), extending the previous work on reference and modification (Nivre and Croft, 2025). An overview of the construction is shown in Table 1, where we outline which syntactic relations are used to annotate different argument phrases across constructions and strategies. To this should be added the annotation of verbal predicates using part-of-speech tags, features and the *aux* relation, and of argument encoding through morphological features and the *case* relation, as described in Section 3.5.

Based on our review of the existing guidelines and annotation practices, we have also proposed some modifications to the guidelines that should be considered for future versions of UD. This includes modified guidelines for transitive clauses in (some) languages with ergative alignment, and for ditransitive clauses generally, as well as a proposal for a new subtyping system, which will make the annotation of nonbasic voice constructions more transparent. As stated in Nivre and Croft (2025), these proposals need to be evaluated also from other perspectives, since UD is designed as “a very subtle compromise between a number of competing criteria” (de Marneffe et al., 2021, p. 302), and the discussion also needs to be informed by a more comprehensive review of the UD framework, covering all major types of constructions and strategies. It is our goal to continue this review in a series of future publications.

## Acknowledgments

This work received support from the CA21167 COST action UniDive and from the Swedish Research Council (grant no. 2022-02909). We thank André Coneglian, Leonie Weissweiler and three anonymous reviewers for comments and suggestions that helped improve the article.

## References

- John Beavers. 2011. On affectedness. *Natural Language and Linguistic Theory*, 29:335–370.
- Bernard Comrie. 1989. *Language Universals and Linguistic Typology*. University of Chicago Press.
- William Croft. 2010. *Verbs: Aspect and Causal Structure*. Oxford University Press.
- William Croft. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology*, 20(2):377–393.
- William Croft. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge University Press.
- William Croft. in press. Comparative concepts. In Hilary Nesi and Petar Milin, editors, *Encyclopedia of Language and Linguistics*. Elsevier.
- Jon P. Dayley. 1989. *Tzutujil Reference Grammar*. University of California Press.
- Roger M. W. Dixon. 1994. *Ergativity*. Cambridge University Press.
- Joseph H. Greenberg. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of Grammar*, pages 73–113. MIT Press.
- John Haiman. 2011. *Cambodian (Khmer)*. John Benjamins.
- Martin Haspelmath. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86:663–687.
- Martin Haspelmath. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology*, 15:535–567.
- Martin Haspelmath. 2015. Transitivity prominence. In Andrej L. Malchukov and Bernard Comrie, editors, *Valency Classes in the World's Languages*, pages 131–147. Mouton de Gruyter.
- Arthur Lorenzi, Peter Ljunglöf, Ben Lyngfelt, Tiago Timponi Torrent, William Croft, Alexander Ziem, Nina Böbel, Linnéa Bäckström, Peter Uhrig, and Ely A. Matos. 2024. MoCCA: A model of comparative concepts for aligning constructions. In *Proceedings of the 20th Joint ACL – ISO Workshop on Interoperable Semantic Annotation*, pages 93–98.
- Andrej L. Malchukov. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber and Helen de Hoop, editors, *Competition and Variation in Natural Languages: The Case for Case*, pages 73–117. Elsevier.
- Marie de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47:255–308.
- Edith A. Moravcsik. 1978. On the distribution of ergative and accusative patterns. *Lingua*, 45:233–279.
- Joakim Nivre. 2025. Constructions and strategies in Universal Dependencies. In *Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)*, pages 419–423.
- Joakim Nivre and William Croft. 2025. Reference and modification in Universal Dependencies. In *Proceedings of the 8th Workshop on Universal Dependencies (UDW)*, pages 1–10.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Dan Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, pages 1659–1666.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Dan Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, pages 4034–4043.
- Elizabeth Patz. 2002. *A Grammar of the Kuku Yalanji Language of North Queensland*. Australian National University.
- Leonard Talmy. 1988. Force dynamics in language and cognition. *Cognitive Science*, 12:49–100.
- Takasu Tsunoda. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics*, 19:389–438.
- Takasu Tsunoda. 1985. Remarks on transitivity. *Journal of Linguistics*, 21:385–396.
- Francis Tyers and Karina Mishchenkova. 2020. Dependency annotation of noun incorporation in polysynthetic languages. In *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*, pages 195–204.
- David John Weber. 1989. *A Grammar of Huallaga (Huánaco) Quechua*. University of California Press.
- Leonie Weissweiler, Nina Böbel, Kirian Guiller, Santiago Herrera, Wesley Scivetti, Arthur Lorenzi, Nurit Melnik, Archana Bhatia, Hinrich Schütze, Lori Levin, Amir Zeldes, Joakim Nivre, William Croft, and Nathan Schneider. 2024. UCxN: Typologically informed annotation of constructions atop Universal Dependencies. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16919–16932.

C. J. Williams. 1980. *A Grammar of Yuwaalaraay*.  
Australian National University.

Christoph H. Wolfart and Janet F. Carroll. 1981. *Meet  
Cree; A Guide to the Cree Language*. University of  
Nebraska Press.