Linguistic Typology Meets Universal Dependencies

William Croft, Dawn Nordquist, Katherine Looney, Michael Regan
Department of Linguistics
University of New Mexico
{wcroft,nordquis,looneyk,reganman}@unm.edu

Abstract

Current work on universal dependency schemes in NLP does not make reference to the extensive typological research on language universals, but could benefit since many principles are shared between the two enterprises. We propose a revision of the syntactic dependencies in the Universal Dependencies scheme (Nivre et al. [16, 17]) based on four principles derived from contemporary typological theory: dependencies should be based primarily on universal construction types over language-specific strategies; syntactic dependency labels should match lexical feature names for the same function; dependencies should be based on the information packaging function of constructions, not lexical semantic types; and dependencies should keep distinct the "ranks" of the functional dependency tree.

1 Introduction

A number of parsing and tagging schemes have recently been proposed that aim to be universal across languages, including the Universal Stanford Dependencies (USD; de Marneffe et al. [8]) and the Google Universal Part-of-Speech Tagset (Petrov et al. [20]). More recently, the Universal Dependencies (UD) initiative (Nivre et al. [16, 17]) has brought together a slightly altered form of the Universal Stanford Dependencies, an extended version of the Google Universal Part-of-Speech Tagset, and a revised subset of the Interset morphological features (Zeman [31]).

Nivre [16] writes, 'Traditionally, research in our [NLP] community has not paid much attention to language typology or linguistic universals'. Unfortunately, none of these proposals make explicit reference to the extensive typological literature on universals based on large-scale, balanced crosslinguistic samples (although this does not mean that typology has not influenced these proposals). As a consequence, these proposed universal schemes offer a mixture of genuinely universal annotations from a typological perspective, and some annotations that are language-specific, even if they are specific to multiple languages. Universals projects in NLP would benefit from taking a more typological perspective (cf. Bender [1]). First, a universal scheme that better reflects typological patterns across all

of the world's languages should be more robust and extendable to other languages for which digital resources may be developed. Second, the Universal Dependencies initiative is actually quite harmonious with many elements of recent typological theory. In fact, many changes in UD version 2 independently converge with proposals in this paper. [1]

In the following sections, we present the four most important principles behind a typological approach to syntactic analysis. From these principles, we develop an annotation scheme for dependencies that reflects a fuller range of crosslinguistic variation and the universals governing the variation, but—we hope—without losing the practical values embodied in the UD initiative. This scheme is currently used in teaching typological syntax to undergraduates. The pedagogical and theoretical goals can be compared to the principles guiding the UD project:[2]

1. UD needs to be satisfactory on linguistic analysis grounds for individual languages.
2. UD needs to be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be suitable for computer parsing with high accuracy.
5. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. We refer to this as seeking a *habitable* design, and it leads us to favor traditional grammar notions and terminology.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...).

Theoretical and pedagogical goals match with all the UD goals, except 4 and 6, which do not apply. We recognize that some revisions to UD in this paper proposed for theoretical and pedagogical goals may conflict with UD goals 4 and 6. In this paper we restrict our attention to typologically motivated revisions to the UD syntactic dependencies; we do not address multiword expressions, fixes to deal with headless constructions in dependency syntax, or lexical tagging.

2 Constructions and strategies

One needs to develop valid comparative concepts in crosslinguistic comparison (Haspelmath [12]). Haspelmath is following a long tradition in typology (Greenberg [9], Keenan and Comrie [13], Stassen [25], Croft [5]). These authors all argue

[1] The first author thanks Joakim Nivre, Chris Manning and Marie-Catherine de Marneffe for the opportunity to discuss UD v2 before its release.

[2] http://www.universaldependencies.org/introduction

in support of defining comparative concepts in semantic/functional terms. Haspelmath, however, also argues for the need for "hybrid" comparative concepts defined in both semantic and formal (morphosyntactic) terms, as long as the formal criteria are crosslinguistically valid.

Two types of "hybrid" comparative concepts are necessary for crosslinguistic comparison, and in fact have been used in typological research since its inception (Croft [6, 7]). The first type of comparative concept, *constructions*, are any morphosyntactic structure in any language used to express a particular meaning or function. For example, a predicate nominal construction is any construction used to predicate the object category of a referent, as in *Ivan is the best dancer*. The second type of comparative concept, *strategies*, are constructions that use specific morphosyntactic devices, that is, specific formal structures, where those formal structures are defined in a crosslinguistically valid way. For example, the English predicate nominal construction uses a copula strategy, where 'copula' is defined in crosslinguistically valid terms as a (usually free) morpheme, distinct from the object concept word, that is part of the predication. Constructions are universal, in fact universal by definition, since human languages are general-purpose communication systems. Strategies are language-specific, or more precisely, not necessarily universal; English uses a copula strategy in its predicate nominal construction, but many languages do not. This leads to the first principle for designing a typologically universal annotation scheme: **a universal annotation scheme should have a classification of constructions as its universal foundational layer**; it should avoid wherever possible annotations dependent on strategies which are not universal.

UD does the right thing typologically in its aim to 'push all the way the design principle of having direct links between content words' (de Marneffe et al. [8]). This is because the use of an independent syntactic unit such as a copula or adposition is not a strategy found in every language. For this reason, the topology of the dependency trees in our proposal largely matches that of UD; the primary differences are in the classification of the dependencies (see Figure 1). Of course, elements in a universal scheme that represent strategies, at least the most commonly occurring strategies, will also be needed. The most common strategies found as function words include `case` (as in UD), classifiers (`clf`, added to UD v2), independent indexation/agreement elements (`idx`) and linkers (`lnk`); these are given crosslinguistically valid definitions in Croft [5]). Coordinating and subordinating conjunctions (`cc` and `mark`, following UD) are also strategies, as opposed to asyndetic coordination and deranking (a morphological strategy; see Stassen [25]). [3]

Adpositions and case features are different strategies for the same construction (in the typological sense), namely the construction that relates an argument dependent to its head. A universal scheme should capture the fact that adpositions and case morphology are two strategies for the same construction. This leads to a

[3] Not all words categorized as function words are strategies, e.g. pronouns and determiners.

	reference	**modification**	**predication**
object	*the sharp **thorns***	*the **bush's** thorns*	*It's **a thorn***
property	***sharpness***	*the **sharp** thorns*	*Those thorns **are sharp***
action	*I said **that** the thorns **scratched** me.* *the **scratching** of the thorns*	*the thorns **that scratched** me* *the thorns **scratching** me*	*The sharp thorns **scratched** me*

Table 1: Packaging of semantic classes as either referring expression (argument), modifier or predicate.

second design principle for typologically universal annotation: **use the same term for morphological and syntactic strategies for the same function**, across dependency labeling and lexical tagging (morphological features and POS). UD uses the same label, *case*, for the syntactic strategy (dependency of a function word) and the morphological strategy (feature of a word form), which conforms to the second principle. But more generally, UD, like traditional grammar, uses different terms for the independent function word dependency and the corresponding morphological category, and sometimes yet another term for the POS tag for the function word. Here UD principles 2 and 5 are in conflict; we opt for 2 while UD opts for 5.

3 Semantics and information packaging

The crosslinguistic analysis of parts of speech is a long-vexed issue in typology as well as other grammatical theories. The solution is to recognize that parts of speech are best defined by the intersection of lexical semantic categories—object, action, property—and information packaging functions—in the case of major parts of speech, the propositional act functions of reference, predication, and modification (Croft [3, 4]). The primary argument for this analysis is that, in principle at least, any semantic class may be expressed through any propositional act, albeit often with a distinct morphosyntactic strategy. This fact is illustrated with the English examples in Table 1.

We argue that this split-level analysis of meaning into semantic content and information packaging applies to all levels of sentence meaning. Predicates have multiple arguments. They are generally packaged into a ranking that represents discourse salience, essentially subject < object < oblique. The ranking has a preferred realization of semantic roles, with agents as subjects, patients or themes as objects, and other roles as obliques. However, almost any semantic role can be expressed as either subject, object or oblique, illustrated with English examples in Table 2. The less common realizations of semantic roles may be realized grammatically by voice, including verb alternations in English and applicative constructions (Peterson [19]) in other languages. Although agent-like participants do not receive object-like marking in English, they do in the inverse voice construction in Al-

	core		oblique
	subject	**object**	
agent	***The protesters*** *sprayed green paint on the sidewalk.*	—	*The sidewalk was sprayed with green paint* ***by the protesters.***
theme	***Green paint*** *was sprayed on the sidewalk.*	*The protesters sprayed* ***green paint*** *on the sidewalk.*	*The protesters sprayed the sidewalk* ***with green paint.***
goal	***The sidewalk*** *was sprayed with green paint.*	*The protesters sprayed* ***the sidewalk*** *with green paint.*	*The protesters sprayed green paint* ***on the sidewalk.***

Table 2: Packaging of semantic roles in events as either subject, object or oblique.

gonkian and other language families, and in the voice systems of the extensive Austronesian family, e.g. Tagalog.

Coordination and adverbial subordination represent two different ways of packaging temporal, causal, conditional, concessive and other semantic relations between two events. Any of these semantic relations can be expressed in English by either adverbial subordination or by coordination; see Table 3. Expressing a relation between events in terms of coordination packages them as a single complex Gestalt, where the two events are construed in a symmetrical fashion (Wierzbicka [30]). Expressing the same relation between events in terms of adverbial subordination packages them in an asymmetrical relation, usually described as figure (main clause) vs. ground (adverbial clause; Talmy [26], Reinhart [21]).

What does this mean in practical terms? The information packaging functions are generally much more isomorphic to syntactic structures than lexical semantic classes: The information packaging functions are also less variable across languages—i.e. more universal—than lexical semantics. The dissociation of lexical semantics and information packaging leads to a third design principle: **the dependencies should primarily express information packaging structure;** semantic content should be primarily annotated via lexical tagsets. This design principle will lead mostly to a simplification of the UD dependency set. First, *det, nummod* and *amod* would be grouped into `mod`, with their differences captured by lexical semantic tags. UD v2 does not merge these, because some users use only the dependencies while others use only the lexical tags; hence retaining redundancy is preferred. Second, *dobj* and *iobj* would be grouped into `obj`. UD v2 keeps the two separate, though *dobj* is renamed *obj*. Double object constructions generally have the two objects in strict order; although some argue that only one of the two objects is a "true" object, in fact there is no crosslinguistic consistency in the syntactic behavior of the two objects (Peterson [19]). Finally, *nsubj* and *nsubjpass* would be grouped into `sbj`, with semantic roles identified by voice or other semantic tags on the verb. If special dependencies are created for passive voice, then one would have to devise special dependencies for applicative, applicative+passive, and other combinations of valency-changing operations. For this reason, UD v2 has eliminated

Semantic relation	Subordination construction	Coordination construction
Anterior	*He washed the car **before** driving to the party.*	*He washed the car **and** drove to the party.*
Posterior	*He drove to the party **after** washing the car.*	*He washed the car **and** drove to the party.*
Overlap	*He washed the car **while** the sun was still shining.*	*The sun was shining **and** he was washing the car.*
Cause	*She went to bed **because** she was exhausted.*	*She was exhausted **and (so)** went to bed.*
Purpose	*I will grab a stick **to** defend myself.*	*I will grab a stick **and** defend myself.*
Apprehensional	*I grabbed a stick **lest** he attack me.*	*Grab a stick **or** he will attack you.*
Means/Positive Circumstantial	*He got into the army **by** lying about his age.*	*He lied about his age **and** got into the army.*
Negative Circumstantial	*She carried the punch into the living room **without** spilling a drop.*	*She carried the punch into the living room, **and** she didn't spill a drop.*
Additive	***In addition to** having your hand stamped, you must show your ticket stub.*	*You have to have your hand stamped **and** show your ticket stub.*
Substitutive	*We barbecued chicken at home **instead of** going out to eat.*	*We didn't go out to eat, **and/but** barbecued chicken at home.*
Subtractive	*He did all the problems correctly **except** he missed the proof on the last one.*	*He did all the problems correctly **but** he missed the proof on the last one.*
Conditional	***If** you do that, the terrorists have won.*	*Murphy, you do that **and** the terrorists have won, ...*
Concessive	***Although** John had no money, he went into this expensive restaurant.*	*John had no money, **but** he went into this expensive restaurant (anyway).*

Table 3: Packaging of semantic relations between events as either coordination or (adverbial) subordination.

Info packaging rank	Proposed dependencies
Complex sentences	`conj, advcl`
Predicate complex	`cxp, sec` [[PROPERTY]], `aux` [[TAMP]]
Arguments	`sbj, obj, obl, comp` [[EVENT]]
Modifiers	`mod, nmod` [[ENTITY]], `acl` [[EVENT]], `appos`
Admodifiers	`qlfy`
Common strategies	`cc, mark, case, clf, idx, lnk`

Table 4: Summary of proposed revised UD syntactic dependencies

the special dependencies for passive voice, retaining the label *nsubj*.

4 "Ranks" of syntactic structure

Finally, we propose a fourth design principle for the categorization or labeling of dependencies: **dependencies should be unique to the "rank" of dependency structure**. The "ranks" of dependency structure are: predicates, which are chained together in complex sentences (`conj, advcl`, as in UD) but also form complex predicates (`cxp`); arguments, divided into subject, object and oblique (`sbj, obj, obl`); modifiers; and admodifiers (modifiers of modifiers; `qlfy`); see Table 4.

One of these "ranks", complex predicates, is not found in Western traditional grammar. However, complex predicates occur widely in English and other languages, and are diverse in form and function. Certain elements of complex predicates are analyzed as compounds in UD v2, including serial verbs, verb-particle constructions, and more grammaticalized light verb constructions. However, these predicate complexes are syntactically flexible, unlike nominal compounds (*door handle; sky blue*) which are semi-fixed in the terminology of Sag et al. [22]. We distinguish complex predicate structure with a dependency distinct from nominal compounds, `cxp`. UD v2 retains a distinct *cop* dependency for copulas, albeit more narrowly defined than in v1. We follow FrameNet in analyzing copulas as instances of support verbs (light verbs) and therefore eliminate *cop*.

We distinguish two special classes of complex predicate elements. The first are property concepts expressed as secondary predication (secondary predicates (depictives and resultatives) and manner "adverbs". UD treats all three differently: depictives as adverbial clauses (*acl*), resultatives as controlled complements (*xcomp*), and manner adverbs as adverbs (*advmod*). Although English and other Western European languages use a different strategy for manner "adverbs" than for depictives and resultatives, typologically the three property concept constructions share the same range of strategies and indeed can share the same strategy within a language (Loeb-Diehl [14], van der Auwera and Malchukov [27], Verkerk [28, 29]). We distinguish this element of a predicate complex as `sec`. Finally, forms expressing tense, aspect, mood, modality, evidentiality and polarity are called `aux`. UD v1 used the *aux* dependency for such elements when they are verblike, which is only one strategy for expressing these categories, and distinguished negative polarity

(*neg*). UD v2 extends *aux* to uninflected elements expressing such meanings and eliminates the *neg* dependency, but uses *advmod* for aspectual, modal etc. adverbs.

Admodifiers are "adverbs" that qualify modifiers, such as degree and hedging. In our proposed revisions, there is no *advmod* dependency. Manner adverbs are subsumed under secondary predicates (`sec`); adverbs expressing aspect, modality and so on are subsumed under `aux`; adverbs expressing spatial or temporal location (*here, yesterday*) are analyzed as pronominal oblique dependents (`obl`); and adverbs functioning as admodifiers, which generally express degree or hedging, are assigned to the `qlfy` dependency. All of these types of "adverbs" are diverse both in terms of syntactic distribution, morphological form and semantics. A single "adverb" dependency is therefore not a coherent category in grammatical or semantic terms.

From a pure dependency syntax perspective, the "ranks" of complex sentences, complex predicates and arguments are not differentiated because all of them are dependents of the predicate. Linguistically, however, they are very different: adverbial clauses are an asymmetric packaging of information otherwise expressed symmetrically in coordination; complex predicates are complex expressions describing events and their semantic properties; arguments are participants in events. Although we retain the dependency structure for these three "ranks" (all dependent on the predicate), we capture these distinctions by using distinct dependencies, as seen in Table 4.

The fourth design principle makes dependency structure most closely match information packaging structure. However, it collides with another fact—the fact that motivates the separation of lexical semantics from information packaging in the first place: in principle, any semantic category can be packaged in any way. To illustrate the practical problem: an event may function as an argument of a predicate (1); but then the event's own participant entities (John, the request) may be expressed as "arguments" of the event "predicate". Of course, languages allow these packagings of content, but they also use a range of strategies to express them. In the example of complements, strategies range from predicate-like tense-aspect-mood inflection and argument-like subject/object/oblique coding (1), to argument-like nominalized event forms and possessive modifier-like coding of the participant entities (2), to—even more problematic from a practical point of view—the mixing in a single construction of strategies from prototypical predicate-argument and argument-modifier constructions (for example, English gerunds; 3).

(1) She believes that John approved the request.

(2) She regrets John's approval of the request.

(3) She regrets John's approving the request.

The solution adopted here for this practical problem is twofold. First, isolate the most common cases of this complexity and give them their own distinct dependency label, based on their semantic categories: events as arguments are complements (`comp`), events as modifiers are adjectival (relative) clauses (`acl`), entities

as modifiers (possessives, etc.) are nominal modifiers (nmod), as in UD. [4] Second, allow for recursion of dependencies with these distinct dependencies. Hence complements and relative clauses may have arguments that are dependents of the subordinate clause predicate, and noun modifiers may themselves have modifiers that are dependents of the noun modifier. These exceptions are justified because they are the cases that most frequently—but by no means always—have complex syntax "imported" from the "native" position of the semantic category in the dependency tree. This is basically a standard analysis of recursive structures, but limited to just these dependencies, and with the parallels to other strategies highlighted (we would analyze 3 and even 2 as comp).

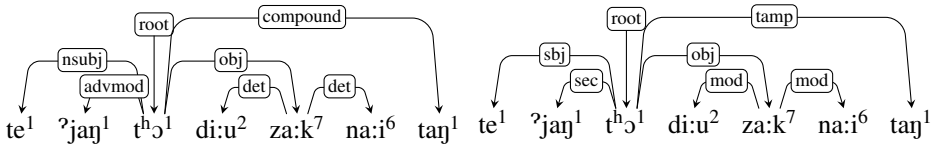
The final scheme for syntactic dependencies that conform to the four typological principles is given in Table 4. UD dependencies left unchanged are: the root (*root*); multiword expressions (*fixed, flat, compound* in UD v2; cf. Sag et al. [22]); discourse (*vocative, discourse*), loose joining (*dislocated, list, parataxis, reparandum*); elliptical elements (*orphan*); unspecified (*dep*); punctuation (*punct*) and other special cases (*reparandum, goeswith*). We are unsure about dividing Sag et al.'s [22] semi-fixed expressions into exocentric (*flat*) and endocentric (*compound*) types, and we exclude complex predicate elements from compound. The discourse and loose joining dependencies may be revised, but at present insufficient typological research has been done on them.

These four principles guided the proposed revisions to UD in this paper, but there are other ways in which typology can provide input to a universal syntactic annotation scheme. For example, sentence coordination is unheaded, and various analyses have been proposed (Zeman et al. [32]). UD v1 attached following conjuncts to the first conjunct. However, two typological phenomena suggest that a chaining analysis is preferable. The first is the class of switch-reference constructions (Haiman and Munro [11]), in which verb forms are selected based on the coreference or lack thereof between the conjunct subject and the subject of the preceding or following conjunct. The second is the universal phenomenon of tense iconicity (Haiman [10]): the sequence of conjoined clauses mirrors the sequence of events. UD v2 adopts the chaining analysis.

5 Applications: examples and annotation task using the proposed scheme

Figure 1 gives the UD v2 and proposed UD dependency trees for example 4, a sentence from a Maonan text (Lu [15]).

[4] UD divides complements into subject (*scomp*), obligatory control (*xcomp*) and nonobligatory control (*ccomp*) complement types. We do not distinguish complements by grammatical role, in order to avoid proliferation of complement types. Control properties are lexically or pragmatically determined; if lexical, then by the third principle they would be lexically tagged. We recognize that retaining the distinction between *ccomp* and *xcomp* may be useful for UD goal 6.



Maonan: UD v2 Maonan: Typological UD

Figure 1: Comparison of UD v2 and typological UD dependency trees.

Language	**Typological traits**	**Cohen's kappa**
Nguna (Austronesian, Oceania)	VO, some affixing	0.834
Chantyal (Sino-Tibetan, S Asia)	OV, some affixing	0.803
Arapaho (Algonkian, N America)	free, polysynthetic	0.849
Supyire (Niger-Congo, W Africa)	VO, some affixing	0.873
Maonan (Daic, E Asia)	VO, highly analytic	0.745
Mapuche (Araucanian, S America)	OV, heavily affixing	0.775
All languages	1126 annotations	0.805

Table 5: Results of annotation task for six languages

(4) te^{1} ʔjaŋ^{1} tʰɔ^{1} di:u^{2} za:k^{7} na:i^{6} taŋ^{1}
3PL slowly drag the.CL rope this come
They slowly started to drag the string.

We performed an annotation task using the proposed scheme. Two annotators, one an instructor of the syntax class, the other an undergraduate who had not used the annotation scheme previously, annotated passages from six languages (Nguna [23], Chantyal [18], Arapaho [text formerly available at the University of Colorado Arapaho website], Maonan [15], Supyire [2] and Mapuche [24]), already divided into 10-20 single sentences. The annotators used a hybrid annotation scheme which includes constituents for modifier phrases, argument phrases, complex predidates (which may be discontinuous) and clauses, as well as labeling of heads and dependent types. The hybrid scheme has proven to be an effective pedagogical tool. The hybrid annotation can be translated into a pure dependency tree (we have not yet automated the translation). However, the hybrid annotation also allows for the annotation of headless and exocentric constructions, where a pure dependency analysis requires conventional assignment of an arbitrary head, and rather complex rules for doing so, as seen in the UD guidelines. Scoring of interannotator agreement is done by an alignment algorithm; this scoring provides an intuitively natural weighting of the different types of inconsistencies or errors.

The results of the annotation task are given in Table 5. Interannotator agreement is measured using Cohen's kappa. Unsurprisingly, higher agreement is found for languages typologically similar to English; Arapaho is high because some sentences are single words in a polysynthetic language.

References

[1] Emily M. Bender. Linguistically naïve != language independent: why NLP needs linguistic typology. In *Proceedings of the EACL 2009 Workshop on the Interaction between Linguistics and Computational Linguistics*, pages 26–32. Association for Computational Linguistics, 2009.

[2] Robert Joel Carlson. *A Grammar of Supyire*. Berlin: Mouton de Gruyter, 1994.

[3] William Croft. *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press, 1991.

[4] William Croft. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press, 2001.

[5] William Croft. *Typology and Universals*. Cambridge: Cambridge University Press, 2nd edition, 2003.

[6] William Croft. Comparing categories and constructions crosslinguistically (again): the diversity of ditransitives (review article on *Studies in Ditransitive Constructions: A Comparative Handbook*, edited by Andrej Malchukov, Martin Haspelmath and Bernard Comrie). *Linguistic Typology*, 18:533–51, 2014.

[7] William Croft. Comparative concepts and language-specific categories: theory and practice. *Linguistic Typology*, 20:377–393, 2016.

[8] Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. Universal Stanford Dependencies: a cross-linguistic typology. In *Proceedings of LREC 2014*, pages 4585–4592, 2014.

[9] Joseph H. Greenberg. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of grammar*, pages 73–113. Cambridge, Mass.: MIT Press, 2nd edition, 1966.

[10] John Haiman. *Natural Syntax: Iconicity and Erosion*. Cambridge: Cambridge University Press, 1985.

[11] John Haiman and Pamela Munro, editors. *Switch-Reference and Universal Grammar*. Amsterdam: John Benjamins, 1983.

[12] Martin Haspelmath. Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86:663–87, 2010.

[13] Edward L. Keenan and Bernard Comrie. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8:63–99, 1977.

[14] Flora Loeb-Diehl. *The typology of manner expressions*. PhD thesis, Radboud Universiteit Nijmegen, 2005.

[15] Tian Qiao Lu. *A Grammar of Maonan*. Boca Raton, Fl.: Universal Publishers, 2008.

[16] Joakim Nivre. Towards a universal grammar for natural language processing. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 3–16. New York: Springer, 2015.

[17] 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 Daniel Zeman. Universal Dependencies v1: a multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 1659–1666. European Language Resources Association, 2016.

[18] Michael Noonan. Chantyal texts. *Himalayan Linguistics Archive*, 2:1–254, 2005.

[19] David A. Peterson. *Applicative Constructions*. Oxford: Oxford University Press, 2007.

[20] Slav Petrov, Dipanjan Das, and Ryan McDonald. A universal part-of-speech tagset. In *Proceedings of LREC 2012*, pages 2089–2096, 2012.

[21] Tanya Reinhart. Principles of gestalt perception in the temporal organization of narrative texts. *Linguistics*, 22:779–809, 1984.

[22] Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. Multiword expressions: a pain in the neck for NLP. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Third International Conference, CICLing 2002, Mexico City, Mexico, February 17-23, 2002: Proceedings*, pages 1–15. New York: Springer, 2002.

[23] Alfred Schütz. *Nguna Texts*. Honolulu: University of Hawaii Press, 1969.

[24] Ineke Smeets. *A Grammar of Mapuche*. Berlin: Mouton de Gruyter, 2008.

[25] Leon Stassen. *Comparison and Universal Grammar*. Oxford: Basil Blackwell, 1985.

[26] Leonard Talmy. Figure and ground in complex sentences. In Joseph H. Greenberg et al., editor, *Universals of Human Language, 4: Syntax*, pages 625–652. Stanford: Stanford University Press, 1978.

[27] Johan van der Auwera and Andrej L. Malchukov. A semantic map for depictive adjectivals. In Nikolaus P. Himmelmann and Eva Schultze-Berndt, editors, *Secondary Predication and Adverbial Modification: The Typology of Depictives*, pages 393–421. Oxford: Oxford University Press, 2005.

[28] Annemarie Verkerk. Secondary predication in a typological context: the encoding of resultatives, depictives and manner predications and their placement within a conceptual space. Master's thesis, Radboud Universiteit Nijmegen, 2009.

[29] Annemarie Verkerk. A semantic map of secondary predication. In *Linguistics in the Netherlands 2009*, pages 115–26. Amsterdam: John Benjamins, 2009.

[30] Anna Wierzbicka. Coordination: the semantics of syntactic constructions. In *Lingua Mentalis: The Semantics of Natural Language*, pages 223–85. New York: Academic Press, 1980.

[31] Daniel Zeman. Reusable tagset conversion using tagset drivers. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, pages 213–218, 2008.

[32] Daniel Zeman, Ondřej Dušek, Martin Popel Mareček, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabortský, and Jan Hajič. HamleDT: Harmonized multi-language dependency treebank. *Language Resources & Evaluation*, 48:601–637, 2014.