Speech Act Constructions in Universal Dependencies

Joakim Nivre, William Croft, André V. Lopes Coneglian
Uppsala University, University of New Mexico, Federal University of Minas Gerais
joakim.nivre@lingfil.uu.se, wacroft@icloud.com, coneglian@ufmg.br

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 speech act constructions. We find that UD currently lack mechanisms for systematically capturing speech act constructions and briefly discuss ways in which this can be remedied.

Keywords: Universal Dependencies, speech act constructions

1. Introduction

Universal Dependencies (UD) is an annotation framework for morphosyntax, designed to be applicable to all human languages and to enable meaningful cross-linguistic comparisons (Nivre et al., 2016, 2020; de Marneffe et al., 2021). To find out whether UD meets these requirements, Nivre (2025) proposes to build a constructicon 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).

In this framework, *constructions* are form-function pairings defined solely in terms of their function (hence universal), while *strategies* are defined by the pairing of a function with some cross-linguistically identifiable morphosyntactic form. For example, the property predication construction is defined as any construction in any language where a property concept is predicated of a reference, as in the English *he is tall*. The verbal copula strategy describes an object predication construction that involves a linking element in the form of a verb, like English *is*. Other common strategies in the world's language are the nonverbal copula strategy (with a nonverbal linking element) and the zero strategy (with no linking element) (Croft, 2022; Nivre et al., 2026).

In order to build a constructicon, we first need to check if the current UD guidelines cover a given construction and all its attested strategies and propose additions when needed. Moreover, in the same process, we can assess to what extent the annotations of different strategies capture the fact that they are realizations of the same construction, which facilitates cross-linguistic comparisons. A further step would be to investigate how actual treebanks have dealt with the phenomena at hand, which is outside the scope of the current investigation.

In this paper, we review how UD handles speech act constructions, which are discussed in Chapter 12 of Croft (2022). This complements previous papers on reference and modification (Nivre and Croft, 2025) and verbal predication (Croft and Nivre, 2025), as well as concurrent papers on nonprototypical predication and nonpredicational clauses (Nivre et al., 2026) and on complex predicates (Coneglian et al., 2026).

2. Speech Act Constructions

Speech acts can in general be defined in terms of the response that a speaker wants or requires from an addressee. In this context, we are interested in grammatical constructions that encode speech act functions. Typological surveys have shown that there are three such constructions found in all or most of the world's languages: the declarative, the interrogative, and the imperative-hortative construction; a fourth construction, the exclamative construction, is less common as a distinct construction but still attested in many languages (Sadock and Zwicky, 1985; König and Siemund, 2007; Croft, 2022).

The *declarative* construction is the default type of speech act construction and simply asserts some propositional content, inviting the addressee to accept it. The *interrogative* construction, or question, requests some information from the addressee, while the *imperative-hortative* construction requests that an action be carried out. The *exclamative* construction, finally, expresses a strong emotional reaction to a propositional content. Since this construction is less commonly attested, and the strategies used to realize it very diverse, we will not discuss it further in this paper.

As already mentioned, the declarative construction can be regarded as the default construction and will therefore not be discussed in this paper, as it is covered in depth in previous papers on clause constructions (Croft and Nivre, 2025; Nivre et al., 2026). Before we turn to the other speech act constructions, however, we need to discuss *negation* in declarative constructions and the related notion of *polarity*.

2.1. Negation of Declaratives

Declarative constructions with positive polarity constitute the default speech act construction and are overwhelmingly zero coded in the world's languages. That is, there is no overt coding over and above the coding of the clause construction itself. Negative polarity, by contrast, is normally overtly coded using one of four strategies, illustrated in examples (1–4).

(1) díayi yiíre baka-rí-ti
dog 1SG-SPEC bite-NEG-EVID
'the dog didn't bite me'

(2) à sə́ɗà cébébè pày
3SG.M know jackal NEG
'he didn't see the jackal'

(3) e-n syö-nyt omena-a
NEG-1SG eat-PTCP apple-PRTT
'I didn't eat an apple'

(4) e sega ni la'o o Jone
3SG NEG COMP go ART John
'John is not going'

In (1), from Tuyuca (Barnes, 1994), negation is expressed as an overt bound morpheme affixed on the verb. In (2), from Musgu (Meyer-Bahlburg, 1972; Dryer, 2005), it is instead realized as an uninflected independent word form. These are the two most common strategies in the world's languages. Less common are the strategies illustrated in (3), from Finnish (Sulkala and Karjalainen, 1992; Dryer, 2005), and in (4), from Boumaa Fijian (Dixon, 1988; Dryer, 2005), where the negation is realized as a verb. In the former case, the negative verb forms a complex predicate together with a participial form of the main verb; in the latter, the main verb is in a subordinate clause introduced by the complementizer *ni*, a strategy that can be paraphrased as 'it is not the case that John is going'.

2.2. Interrogatives

Interrogative constructions are commonly divided into three types depending on function: polarity questions (sometimes called yes/no-questions), information questions (sometimes called wh-questions), and alternative questions. We will focus on polarity questions and information questions, which have been the most studied typologically.

For polarity questions, the two most common strategies are a distinct intonation contour (almost always a rising one) and/or an overt interrogative particle (which may be realized as an independent word or as an affix on the verb). The use of a distinct word order pattern, familiar from Germanic languages, is a much less common strategy.

Information questions typically contain an interrogative pronoun or modifier for the entity to be identified in the answer to the question, and the use of a distinct intonation contour and/or an interrogative particle is much less common in this case. In a large minority of languages, the interrogative pronoun obligatorily occurs at the beginning of the sentence, but it is more common that it occurs in the position where the corresponding element would occur.

2.3. Imperative-Hortatives

Imperative-hortative constructions are a wider class than the traditional notion of imperatives and include cases like the following, where there is an appeal to the addressee to help

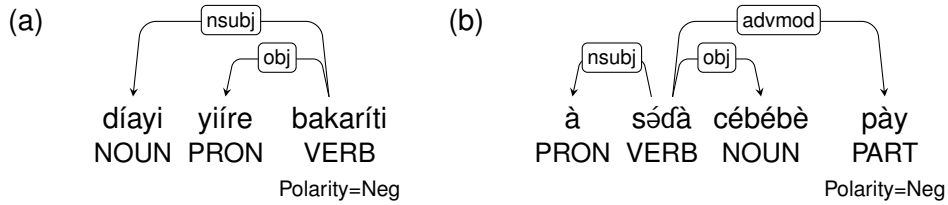

Figure 1: Simplified UD annotation of different strategies for negation of declaratives: (a) affix, (b) word.

bring about an event even though the event is not always an action specifically carried out by the addressee:

(5) a. sing
b. let's sing
c. let her sing
d. let the party begin

The strategies used to realize imperative-hortative constructions are extremely varied and dependent on factors like person, politeness and social relations. The most distinctive strategy used, particularly in the second person, is reduction in form, and it is common to use the bare verb stem as in (5a) above.

3. UD Annotation

Let us begin by observing that the annotation of speech act types is essentially a blank spot in the UD framework. It is not discussed explicitly in the guidelines, and there is no mechanism for annotating this kind of information systematically. This can partly be seen as a consequence of the word-based nature of the UD annotation, which means that linguistic information has to be annotated either as properties of an individual word, through part-of-speech tags and morphological features, or as relations between pairs of words, through typed syntactic relations. Speech act types, however, are properties of entire clauses or sentences and often cannot be associated with individual words or pairs of words.

Technically speaking, given the hierarchical nature of the syntactic annotation, it is possible to annotate information about a larger unit on the head word of that unit (typically the predicate in the case of a clause). However, the scope of such an annotation will in principle always be ambiguous, since there is no mechanism to indicate whether it applies only to the head word or to a larger segment (and if so, which segment). The idea of introducing special so-called "phrase level features" has been discussed on and off in the UD community, but so far no such concept has been introduced.

In the current UD guidelines, there are hence no morphological features or syntactic relations devoted specifically to the annotation of speech act constructions. However, information about such constructions is nevertheless available in a few cases where the strategy for realizing the construction involves a word or affix carrying a distinctive feature. In the following subsections, we will review the most important cases where this happens.

3.1. Negation of Declaratives

In version 1 of UD (Nivre et al., 2016), there was a syntactic relation *neg* reserved for negation, but not restricted to negation at the clause level. Hence, while the negation *not* would be attached with the *neg* relation to the head verb *know* in a declarative cluase like *she does not know*, the same relation would be used to link the determiner *no* to the head noun *idea* in a nominal like *no idea*. The *neg* relation has been deprecated in version 2 of UD (Nivre et al., 2020) because it was deemed to encode a semantic property rather than a morphosyntactic relation.

In the current UD guidelines, the only way to annotate negative polarity is by using the morphological feature Polarity=Neg. This feature can be added both to words containing negation affixes, normally a verb or an auxiliary, and to words expressing negation by themselves. The former strategy is illustrated in Figure 1(a), which shows the UD annotation of example (1);

the latter strategy is illustrated in Figure 1(b), which is the UD annotation of example (2).[1] In many cases, it is therefore possible to determine whether a clause is negated by checking for the occurrence of the feature Polarity=Neg either on the main predicate of the clause or on a direct dependent of the predicate (usually with the relation *advmod*). However, such a procedure is only heuristic and will give the wrong result for sentences like *they did not only dance, they sang as well*, where the word *not* is attached to the verb *dance* and carries the feature Polarity=Neg but still does not negate the clause. For further discussion of negation in UD, see Findlay and Haug (2025).

3.2. Interrogatives

Whether interrogative constructions are captured in the UD annotation depends both on the type of question and on the strategy used to realize it. Starting with polarity questions, the prosody strategy is not captured at all, as UD does not annotate prosodic features, and neither is the less common word order strategy. The question particle strategy should in principle be within reach, since a morphological feature could be assigned to the particle, regardless of whether it is realized as an affix or as an independent word. However, even though the guidelines mention that question particles can be assigned the part-of-speech tag PART, there is currently no morphological feature available for classifying question particles. The only related feature is Mood=Int, which can be used for verbs that have a special form in interrogative clauses, a strategy attested in Turkic languages.

Moving on to information questions, these are generally captured more straightforwardly because they contain an interrogative pronoun, determiner or adverb that carries the feature PronType=Int. One complication is that the word carrying this feature may not be directly attached to the root of the clause. For example, in a sentence like *which student's mother's book are you looking for*, the interrogative determiner *which* is at a distance of four dependency arcs from the verb *looking*. In addition, we need to distinguish direct questions from embedded interrogative clauses, which is in most cases possible by inspecting the position of the clause in the dependency tree.

3.3. Imperative-Hortatives

Of the many diverse strategies employed for imperative-hortative constructions, only those that involve verbal morphology are currently captured in UD through the morphological feature Mood. In addition to the very common value Mood=Imp, for imperative verb forms, the value Mood=Jus can be used for the jussive, which is found, for example, in Arabic.

4. Discussion

Speech act constructions like declaratives, interrogatives and imperative-hortatives, sometimes referred to as basic sentence types, are among the most general linguistic constructions and should arguably be captured in any framework for morphosyntactic annotation. It is noteworthy that UD currently lacks mechanisms for a systematic annotation of these constructions, and it is highly relevant to discuss ways in which this could be remedied.

In the short term, within version 2 of UD, the obvious solution is to add new features and feature values to describe at least the basic speech act constructions declarative, interrogative, imperative-hortative, and exclamative, as well as clause-level negation. For interrogative and imperative-hortative constructions, it may also be useful to distinguish subtypes. It is true that, in a word-based system like UD, the scope of these features will be ambiguous, but this seems preferable to not including the information at all.

In addition to adding new features, one may consider modifying the syntactic relations used to annotate strategies that involve a special word order. For example, in information questions where the interrogative pronoun or adverb obligatorily occurs at the beginning of the sentence, the *dislocated* relation could be used with a special subtype to capture the word order strategy.

[1] In these examples, we simplify the UD representations by omitting (a) lemmas and (b) morphological features that are not relevant for the discussion. For reasons of space, we do not repeat glosses and translations in the examples.

In the longer term, in subsequent versions of UD, it may be worth introducing a mechanism for using features that apply to larger units than words. Such a mechanism would be useful not only for the annotation of speech act constructions, but more generally for categories that transcend word boundaries, for example, complex predicates.

5. Conclusion

In this paper, we have taken another step towards a constructicon for UD, in the sense of Nivre (2025), by reviewing the way UD annotates speech act constructions, following the taxonomy of Croft (2022), thereby extending the previous work on reference and modification (Nivre and Croft, 2025) and verbal predication (Croft and Nivre, 2025).[2] The main findings in this paper are negative, since UD currently lack mechanisms for systematically capturing speech act constructions. We have briefly discussed how this can be remedied both in the long and the short term.

6. Acknowledgments

This work received support from the CA21167 COST action UniDive, funded by COST (European Cooperation in Science and Technology), from the Swedish Research Council (grant no. 2022-02909), and from the Brazilian National Council for Scientific and Tecnhnological Development, CNPq (grant no. 153972/2025-4 and no. 405484/2025-9). We thank Leonie Weissweiler and three anonymous reviewers for comments and suggestions that helped improve the article.

7. Bibliographic References

Janet Barnes. 1994. Tuyuca. In Peter Kahrel and René van den Berg, editors, *Typological Studies in Negation*, pages 325–342. John Benjamins.

André V. Lopes Coneglian, Joakim Nivre, and William Croft. 2026. Complex predicates in Universal Dependencies. In *Proceedings of the Ninth Workshop on Universal Dependencies*.

William Croft. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge University Press.

William Croft and Joakim Nivre. 2025. Verbal predication constructions in Universal Dependencies. In *Proceedings of the Second International Workshop on Construction Grammars and NLP*, pages 50–60.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47:255–308.

Robert M. W. Dixon. 1988. *A Grammar of Boumaa Fijian*. University of Chicago Press.

Matthew S. Dryer. 2005. Negative morphemes. In Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, editors, *World Atlas of Language Structures*, pages 454–457. Oxford University Press.

Jamie Yates Findlay and Dag Trygve Truslew Haug. 2025. Negation in Universal Dependencies. In *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)*.

Ekkehart König and Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shope, editor, *Language Typology and Syntactic Description, 2nd Edition. Volume 1: Clause Structure*, pages 276–324. Cambridge University Press.

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 constructicons. In *Proceedings of the 20th Joint ACL – ISO Workshop on Interoperable Semantic Annotation*, pages 93–98.

Hilke Meyer-Bahlburg. 1972. *Studien zur Morphologie und Syntax des Musgu*. Helmut Buske.

[2] In concurrent papers, we review nonprototypical predication and nonpredicational clauses (Nivre et al., 2026) and complex predicates (Coneglian et al., 2026).

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, William Croft, and André Coneglian. 2026. Nonprototypical predication and nonpredicational clauses in Universal Dependencies. In *Proceedings of the Ninth Workshop on Universal Dependencies*.

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.

Jerrold M. Sadock and A. Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shope, editor, *Language Typology and Syntactic Description. Volume 1: Clause Structure*, pages 155–196. Cambridge University Press.

Helen Sulkala and Merja Karjalainen. 1992. *Finnish*. Routledge.