Gapping Constructions in Universal Dependencies v2

Sebastian Schuster **Matthew Lamm** **Christopher D. Manning**
Department of Linguistics and Department of Computer Science
Stanford University, Stanford, CA 94305
{sebschu,mlamm,manning}@stanford.edu

Abstract

In this paper, we provide a detailed account of sentences with gapping such as "*John likes tea, and Mary coffee*" within the Universal Dependencies (UD) framework. We explain how common gapping constructions as well as rare complex constructions can be analyzed on the basis of examples in Dutch, English, Farsi, German, Hindi, Japanese, and Turkish. We further argue why the adopted analysis of these constructions in UD version 2 is better suited for annotating treebanks and parsing than previous proposals, and we discuss how gapping constructions can be analyzed in the *enhanced* UD representation, a graph-based semantic representation for shallow computational semantics tasks.

1 Introduction

An important property of natural languages is that speakers can sometimes omit redundant material. One example of this phenomenon is so-called gapping constructions (Ross, 1970). In such constructions, speakers elide a previously mentioned verb that takes multiple arguments, which leaves behind a clause without its main predicate. For example, in the sentence "*John likes tea, and Mary coffee*", the verb *likes* was elided from the second conjunct.

Sentences with gapping pose practical as well as theoretical challenges to natural language processing tasks. From a practical point of view, it is challenging for natural language processing systems to resolve the gaps, which is necessary to interpret these sentences and extract information from them. Further, these sentences are hard for statistical parsers to parse as part of their structure deviates significantly from canonical clause structures.

From a more theoretical point of view, these constructions pose challenges to designers of dependency representations. Most dependency representations that are used in natural language processing systems (e.g., the adaptation of Mel'čuk (1988) for the CoNLL-08 shared task (Surdeanu et al., 2008); de Marneffe et al. (2006); Nivre et al. (2016)) are concerned with providing surface syntax descriptions without stipulating any additional transformations or empty nodes. Further, virtually all dependency representations consider a verb (either the inflected or the main verb) to be the head of a clause. Consequently, the verb governs all its arguments and modifiers. For these reasons, it is challenging to find a good representation of clauses in which a verb that has multiple dependents was elided, because it is not obvious where and how the remaining dependents should be attached in these cases.

In recent years, the Universal Dependencies (UD) representation (Nivre et al., 2016) has become the dominant dependency representation for annotating treebanks in a large variety of languages. The goal of the UD project is to provide guidelines for cross-linguistically consistent treebank annotations for as many languages as possible. Considering that gapping constructions appear in many languages, these guidelines necessarily also have to include guidelines on how to analyze gapping constructions. While the official guidelines[1] provide basic instructions for the analysis of gapping constructions, they lack a detailed discussion of cross-linguistically attested gapping constructions and a thorough explanation why the adopted guidelines should be preferred over other proposals. The purpose of the present paper is therefore to discuss in detail how different gapping constructions can be analyzed in a variety of lan-

[1] See http://universaldependencies.org/u/overview/specific-syntax.html#ellipsis. The first and the last author were both involved in developing these guidelines.

Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017), pages 123–132,
Gothenburg, Sweden, 22 May 2017.

guages and to provide a theoretical comparison of different proposals based on these examples. We further discuss how gapping constructions should be represented in the *enhanced* Universal Dependencies representation, which aims to be a better representation for shallow natural language understanding tasks and how these design choices can potentially help in downstream tasks.

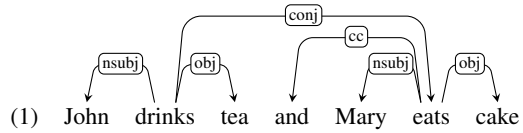
For the purpose of this paper, we consider constructions in which a verb that has multiple dependents was elided, including classic cases of gapping (Ross, 1970). Throughout this paper, we call the elided material (a verb and occasionally also some of its arguments) the GAP. Further, we refer to the dependents of the gap as ORPHANS or REMNANTS, and we refer to the dependents of the verb in the clause with the overt verb as the CORRESPONDENTS, as illustrated with the following annotated sentence.

John	likes	tea	and	Mary	coffee
CORRE-SPONDENT	OVERT VERB	CORRE-SPONDENT		ORPHAN/ REMNANT	ORPHAN/ REMNANT

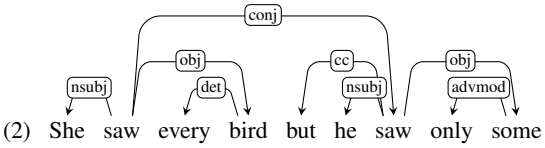
2 Coordination, ellipsis, and gapping in UD v2

Before we discuss how gapping constructions are analyzed in UD v2, we give a brief overview how UD analyzes coordinated clauses and other forms of elliptical constructions.

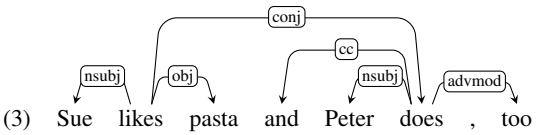
Coordinated clauses are analyzed like all other types of coordination: By convention, the head of the first conjunct is always the head of the coordinated construction and all other conjuncts are attached to the head of the first conjunct with a `conj` relation. If there is an overt coordinating conjunction, it is attached to the head of the succeeding conjunct. This captures the fact that the coordinating conjunction forms a syntactic unit with the succeeding conjunct (Gerdes and Kahane, 2015). A sentence with two coordinated clauses is then analyzed as follows.



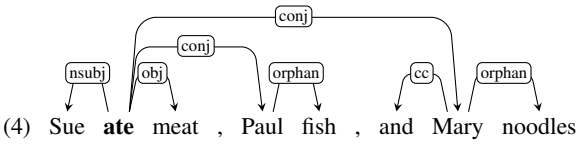
For constructions in which a head nominal was elided, UD promotes the highest dependent according to the hierarchy `amod` > `nummod` > `det` > `nmod` > `case`. The promoted dependent is attached to the governor of the elided nominal with the same relation that would have been used if the nominal had not been elided. All the other dependents of the elided noun are attached to the promoted dependent with their regular relations. For example, in the second conjunct of the following sentence, the head noun *bird* was elided. We therefore promote the determiner *some* to serve as the object of *saw*.



In some cases of ellipsis, a verb phrase is elided but there is still an overt copula or auxiliary verb. In these cases, we promote the copula or auxiliary verb to be the head of the clause and attach all orphans to the auxiliary.



For the constructions we are mainly concerned with in this paper, i.e., gapping constructions in which the governor of multiple phrases was elided, UD v2 adopts a modified version of a proposal by Gerdes and Kahane (2015). We promote the orphan whose grammatical role dominates all other orphans according to an adaptation of the obliqueness hierarchy,[2] to be the head of the conjunct. The motivation behind using such a hierarchy instead of a simpler strategy such as promoting the leftmost phrase is that it leads to a more parallel analysis across languages that differ in word order. We attach all other orphans except for coordinating conjunctions using the special `orphan` relation. Coordinating conjunctions are attached to the head of the following conjunct with the `cc` relation. This leads to the analysis in (4) of a sentence with three conjuncts of which two contain a gap.



[2]Our adaptation prioritizes phrasal over clausal dependents. Translated to UD relations, our adaptation of the obliqueness hierarchy is as follows: `nsubj` > `obj` > `iobj` > `obl` > `advmod` > `csubj` > `xcomp` > `ccomp` > `advcl`. See, for example, Pollard and Sag (1994) for a motivation behind this ordering.

The motivation behind using a special `orphan` relation is that it indicates that the clause contains a gap. If we used a regular relation, it might not be clear that a predicate was elided. For example, if instead, we attached the orphaned subject to the orphaned object using an `nsubj` relation, one could confuse gapping constructions with copular constructions, especially in languages with zero-copula.

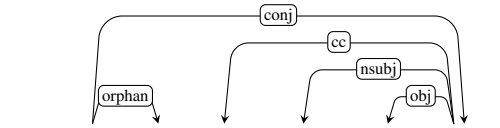
In the rest of this paper, we argue in favor of this proposal for several reasons. First, as we show in the following section, it can be used to analyze a wide range of gapping constructions in many different languages. Second, as we argue in Section 4, there is evidence that the conjunct with the gap forms a syntactic unit, and this fact is captured by the adopted analysis. Finally, as discussed in Section 6, this representation is potentially better suited for automatic parsing than previous proposals.

3 Gapping constructions

We now discuss how a range of attested gapping constructions in a variety of languages can be analyzed according to the above proposal.

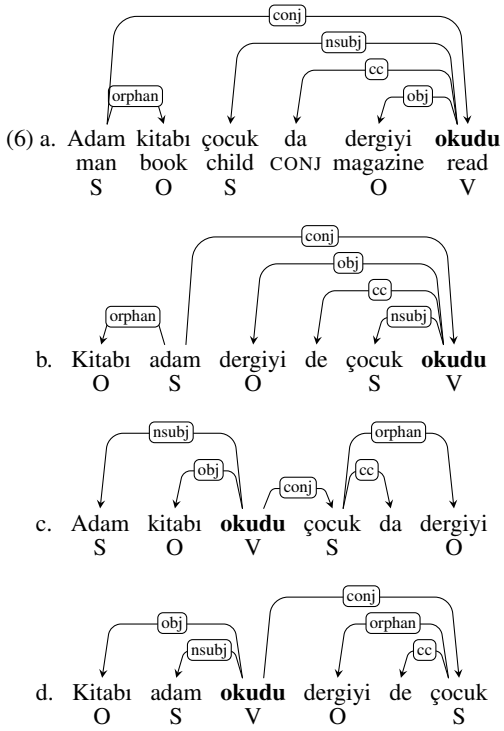
3.1 Single verbs

The most common form of gapping constructions are two or more conjoined clauses in which a single inflected verb is missing in all but one of the conjuncts. As illustrated in (4), in SVO languages such as English, the overt verb typically appears in the first conjunct and is elided from all subsequent conjuncts. In languages with other word orders, the overt verb can also appear exclusively in the last conjunct. For example, in the following sentence in Japanese (an SOV language), the verb appears in the last conjunct and the gap in the first conjunct.



'John bought books, and Mary flowers.' (Kato, 2006)

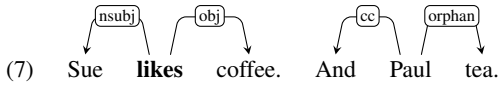
In some languages with flexible word orders such as Turkish, the overt verb can appear in the first or the last conjunct. The orphans typically appear in the same order as the correspondents in the other conjunct as in (6a-d) (Bozsahin, 2000).



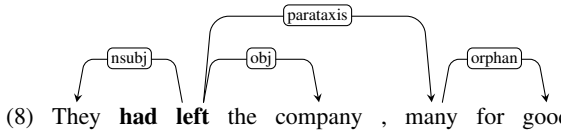
'The man read the book, and the child the magazine.' (Bozsahin, 2000)

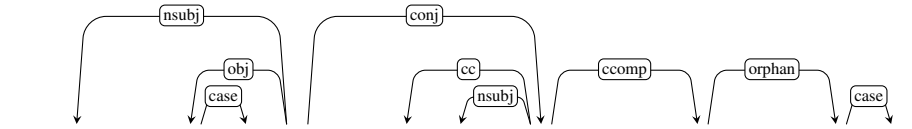
As mentioned above, in order to achieve higher cross-linguistic consistency, the first conjunct is always the head of a coordinated structure in UD. Therefore, the head of the first conjunct is always the head of the coordination, but the internal structure of each conjunct is the same for all four variants in (6).

In some cases, e.g., in certain discourse settings, the clause with the gap is not part of the same sentence as the clause with the overt verb (Gerdes and Kahane, 2015). In these cases, we promote one of the orphans to be the root of the second sentence; the internal structure of the two clauses is the same as when they are part of an intra-sentential coordination.



Further, conjuncts with a gap can also contain additional types of arguments or modifiers. For example, in the sentence in (8), the oblique modifier *for good* does not correspond to any phrase in the first conjunct.



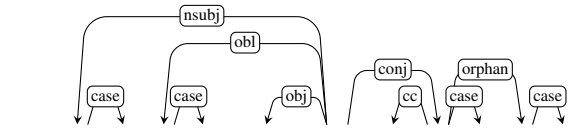


'Mahsa likes this book and Minu knows that her mother (likes) that book' (Farudi, 2013)

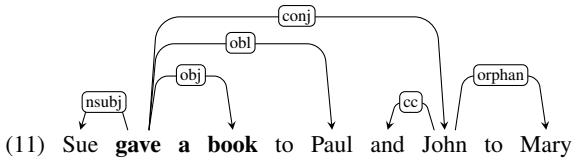
Figure 1: Basic UD tree of a Farsi sentence with a gap within an embedded clause.

3.2 Verbs and their arguments or modifiers

Many languages also allow gapping of verbs along with their arguments or modifiers as illustrated in the following two examples in Hindi (10) and English (11).



'Manu gave a book to Pari and Tanu to Vimla' (Kush, 2016)



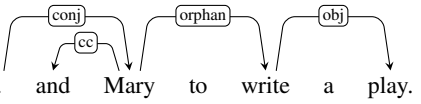
We analyze these cases as analogous to sentences in which only a verb was elided. The subject is promoted to be the head of the second conjunct and the oblique argument is attached with an `orphan` dependency.

3.3 Verbs and clausal complements

Ross (1970) points out that gaps can also correspond to a finite verb and one or more embedded verbs. For example, in the following sentence, it is possible to elide the matrix verb and all or some of the embedded verbs.

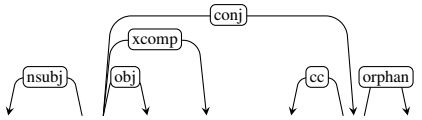
(12) I want to try to begin to write a novel, ...
  a. ... and Mary to try to begin to write a play.
  b. ... and Mary to begin to write a play.
  c. ... and Mary to write a play.
  d. ... and Mary a play. Ross (1970)

In all of these variants, the matrix verb was elided from the second conjunct. While this is an example of subject control and therefore *Mary* is also the subject of all the embedded verbs, it would be misleading to attach *Mary* to one of the embedded verbs because this would hide the fact that the matrix verb was elided. For this reason, we treat *Mary* as the head of the second conjunct and attach the remainder of the embedded clause with an `orphan` relation.

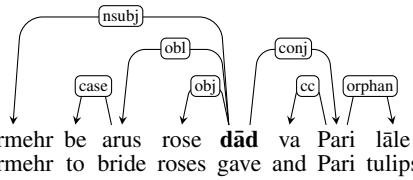


3.4 Non-contiguous gaps

In the previous examples, the gap corresponds to a contiguous sequence in the first conjunct. However, as highlighted by the following examples, this is not always the case.



(adapted from an example in Jackendoff (1971))

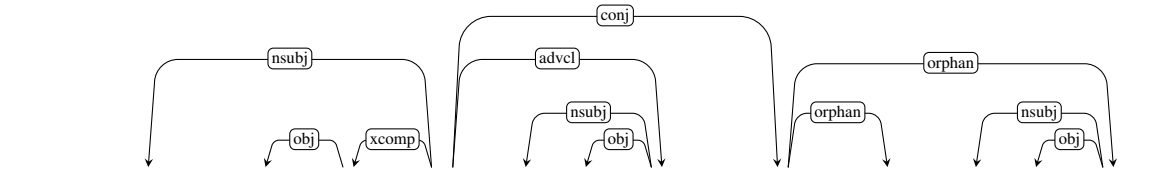


'Farmehr gave roses to the bride and Pari (gave) tulips (to the bride).' (Farudi, 2013)

While the interpretation of the second conjunct is only possible if one fills both gaps, we are also in these cases primarily concerned with the elided verb because neither of the phrases in the second conjunct depend on the second gap. We can therefore analyze constructions with non-contiguous gaps in a similar manner as constructions with contiguous gaps, namely by promoting one orphan to be the head of the conjunct and attaching all other orphans to this head.

3.5 Gaps in embedded clauses

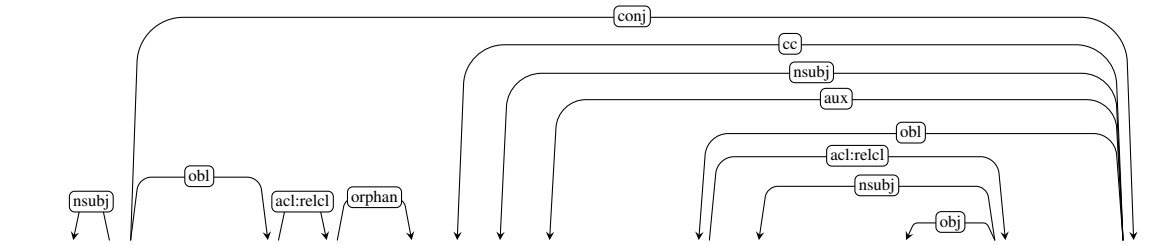
Farudi (2013) notes that in Farsi, gaps can appear in embedded clauses even if the corresponding verb in the first conjunct is not part of an embedded clause. For example, in (9) in Figure 1,



(16) weil P. seinen Freund **besuchen wollte** , was mich beruhigte , und J. seine Kinder , was mich amüsierte
because P. his friend visit wanted , which me reassured , and J. his children , which me amused

"because Peter wanted to visit his friend which reassured me, and Johann (wanted to visit) his children, which amused me"
(Wyngaerd, 2007)

Figure 2: Basic UD tree of a German subordinate clause with an adverbial clause modifying a gap.



(17) Jan heeft met het meisje dat de rode en Piet heeft met de jongen die de witte **wijn binnenbracht gesproken**
Jan has with the girl who the red and Piet has with the boy who the white wine in-brought talked

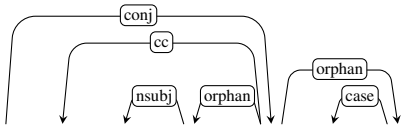
'Jan (talked) to the girl who (brought in) the red (wine), and Piet talked to the boy who brought in the white wine.'
(Wyngaerd, 2007)

Figure 3: Basic UD tree of a Dutch sentence with a gap in a relative clause.

dust ('like') is the main verb of the highest clause of the first conjunct but in the second conjunct, the verb was elided from a clause embedded under *mi-dun-e* ('think'). In these cases, we consider the matrix verb to be the head of the second conjunct as we would if there was no gap, and we promote the subject of the embedded clause (*māmān-esh*) to be the head of the embedded clause. We attach the remaining orphans to the subject with the `orphan` relation.

Note that this construction is different from constructions in which parenthetical material (Pollard and Sag, 1994) appears in the second conjunct, as in the following English example.

(18) [...] I always had a pretty deep emotional connection to him, and I think he to me.[3]



(19) ... and I think he to me

In these cases, we promote the subject of the second conjunct (*he*) and attach the parenthetical *I think* as well as *to him* to the subject.

[3]Source: hhttp://www.ttbook.org/book/transcript/transcript-humour-healing-marc-maron

3.6 Relative clauses

Several Germanic languages such as German and Dutch show more complex gapping behaviors in sentences with adverbial and relative clauses. For example, Wyngaerd (2007) points out that German also allows a verbal gap in clauses modified by an adverbial clause such as the one in Figure 2. In this example, the two verbs *besuchen wollte* ('wanted to visit') are missing from the second clause, which leaves three orphans, namely a subject, a direct object, and an adverbial clause without a governor. As in the case of two orphaned constituents, we promote the subject to be the head of the clause as it is the highest type of argument in the obliqueness hierarchy, and we attach the two other constituents to the subject with an `orphan` relation.

Dutch even allows gaps to appear within relative clauses that modify a constituent in each of the conjuncts. In the example sentence in Figure 3, there are in total two elided verbs, and one elided noun. First, the left conjunct is missing the main verb *gesproken* ('talked') in its matrix clause; second, the relative clause of the object in the first conjunct is missing its verb *binnenbracht* ('brought in'); and third, the noun *wijn* ('wine') was elided from the object in the relative clause. The matrix clause of the first conjunct still con-

tains an auxiliary which we promote to be the head of the first conjunct. We further promote the subject of the relative clause, i.e., the relative pronoun, to be the head of the relative clause, and we attach the adjective, which modifies the elided noun, to the promoted subject with an `orphan` relation.

4 Dependency structure

In the previous section, we showed that the adopted *orphan* analysis can be used to consistently annotate a large number of attested gapping constructions in different languages, which is an important consideration in deciding on an analysis. A second important question is whether our adopted analysis leads to sensible tree structures. Our analysis indicates that a conjunct with a gap forms a syntactic unit, which raises the question whether there is evidence for such a structure.

Many constituency tests such as topicalization, clefting, and stripping suggest that conjuncts with a gap often do not qualify as a constituent. For example, Osborne (2006a) argues against treating the gapping in (20) as the coordination of *[the dog a bone]* and *[the man a flower]*, which would suggest that both conjuncts are constituents. He bases his argument on the observation that the former conjunct fails most constituency tests when it is used in a sentence without coordination (21).

(20) She gave the dog a bone, and the man a flower.

(21) a. *The dog a bone, she gave. (Topicalization)
b. *It was the dog a bone that she gave. (Clefting)
c. ?She gave a dog a bone, not a cat some fish. (Stripping)

However, this argument is based on the assumption that *[the dog a bone]* and *[the man a flower]* form a coordinate structure. If we assume instead that the second conjunct is a clause with elided nodes, then none of the above tests seem applicable. At the same time, as already mentioned above, Gerdes and Kahane (2015) point out that phrases such as *"and the man a flower "* can be uttered by a speaker in response to someone else uttering a phrase such as *"she gave the dog a bone"*. They take this behavior as evidence for treating the entire conjunct with a gap (including the conjunction) as a syntactic unit.

Such an analysis is also in line with most accounts of gapping in the generative literature. While there is disagreement on what the deep structure of sentences with gapping should look like and what transformations are employed to derive the surface structure, there is broad consensus that all remnants are part of the same phrase (e.g., Coppock (2001); Johnson (2009)). We take all of these facts as weak evidence for treating conjuncts with gaps as syntactic units.

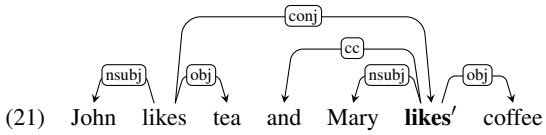
This argument based on constituency criteria might seem surprising considering that such evidence was dismissed when deciding on analyses for other constructions in UD, such as prepositional phrases. One of the major criticisms of UD has been that we attach prepositions to their complement instead of treating them as heads of prepositional phrases because this decision appears to be misguided when one considers constituency tests (see, e.g., Osborne (2015)). However, this decision should not be interpreted as UD completely ignoring constituency. It is true that following Tesnière (1959), UD treats content words with their function words as dissociated nuclei and thus ignores the results of constituency tests for determining the attachment of function words – an approach that is also taken by some generative grammarians, for example, in the form of the notion of extended projection by Grimshaw (1997). But importantly, UD still respects the constituency of nominals, clauses and other larger units. For this reason, it is important to have an analysis of gapping that respects larger constituent boundaries as it is the case with the adopted "orphan" analysis.

5 Enhanced representation

One of the drawbacks of the adopted analysis is that the `orphan` dependencies do not encode information on the type of argument of each remnant, which complicates extracting relations between content words in downstream tasks. However, UD also defines an *enhanced* representation, which may be a graph instead of a tree and which may contain additional nodes and relations (Nivre et al., 2016; Schuster and Manning, 2016). The purpose of this representation is to make implicit relations between words more explicit in order to facilitate shallow natural language understanding tasks such as relation extraction. One property of the *enhanced* representation is that it resolves gaps by adding nodes to *basic* UD trees. Remnants attach to these additional nodes with meaningful re-

lations just as if nothing had been elided,[4] thus solving the issue of the uninformative `orphan` dependencies. The general idea is to insert as many nodes as required to obtain a structure without orphans while keeping the number of additional nodes to a minimum. On top of additional nodes, we add relations between new nodes and existing content words so that there exist explicit relations between each verb and its arguments and modifiers. We now illustrate how different cases of gapping can be analyzed in the *enhanced* representation based on the following representative examples

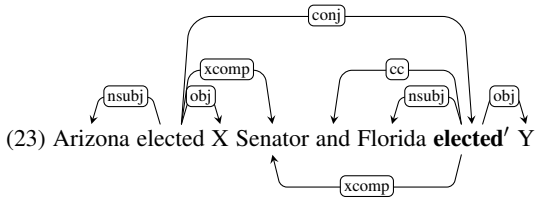
The simplest cases are constructions in which a single verb was elided. In these cases, we insert a copy node[5] of the elided verb at the position of the gap, make this node the head of the conjunct, and attach all orphans to this copy node. For example, for the following sentence, we insert the copy node *likes′* and attach *Mary* as a subject and *coffee* as an object.



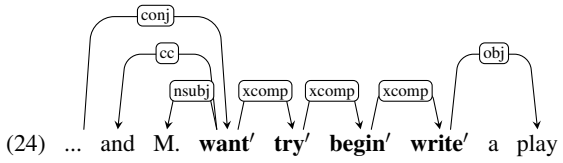
Similarly, we insert a copy node as the new root of a sentence in cases in which the leftmost conjunct contains a gap as, for example, in (5).

'John bought books, and Mary bought flowers.'
(adapted from Kato (2006))

In cases in which arguments or modifiers were elided along with the verb, we still only insert one copy node for the main verb. However, in order to make the relation between the verb and all of its arguments explicit, we also add relations between the new copy node and existing arguments and meaningful modifiers. In (23), we add a copy node for the elided verb *elected* and a relation between the copy node and *Senator*.



In cases in which a finite verb was elided along with one or more embedded verbs, as in the sentence *"I want to try to begin to write a novel, and Mary a play."*, we insert one copy node for each elided verb. However, unlike in the previous example, we do not add relations between the copy nodes and the semantically vacuous function word *to* because it is not required for the interpretation of the sentence.



The motivation behind these design choices is to have direct and meaningful relations between content words. Many shallow natural language understanding systems, which make use of UD such as open relation extraction systems (Mausam et al., 2012; Angeli et al., 2015) or semantic parsers (Andreas et al., 2016; Reddy et al., 2017), use dependency graph patterns to extract information from sentences. These patterns are typically designed for prototypical clause structures, and by augmenting the dependency graph as described above, many patterns that were designed for canonical clause structures also produce the correct results when applied to sentences with gapping constructions.

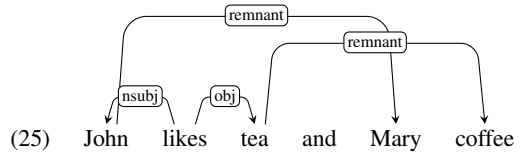
6 Comparison to other proposals

6.1 Remnant analysis

The first version of the UD guidelines (Nivre et al., 2016) proposed that orphans should be attached to their correspondents with the special `remnant` relation. This proposal is very similar to the analysis of string coordination by Osborne (2006b), which adds special orthogonal connections between orphans and correspondents. According to the "remnant" proposal, a sentence with a single verb gap is analyzed as follows.

[4] A similar analysis was used in the tectogrammatical layer of the Prague Dependency Treebank (Bejček et al., 2013).

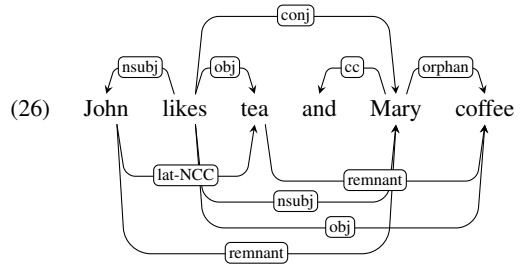
[5] Similar copy nodes are already used for some cases of reduced conjunctions in the *collapsed* and *CCprocessed* Stanford Dependencies representations (de Marneffe and Manning, 2008) and in the *enhanced* UD representation (Schuster and Manning, 2016).



This is arguably a more expressive analysis than the "orphan" analysis because there is a direct link between each orphan and its correspondent, and one is able to determine the type of argument of each orphan by considering the type of argument of its correspondent. However, this analysis comes with several problems. First, it makes it impossible to analyze sentences with orphans that do not have a correspondent such as the sentence with an additional modifier in (8), or sentences whose correspondents appear in a previous sentence as in (7). Second, the `remnant` relations appear to be an abuse of dependency links as they are clearly not true syntactic dependency relations but rather a kind of co-indexing relation between orphans and correspondents. Further, such an analysis introduces many long-distance dependencies and many non-projective dependencies, both of which are known to lower parsing performance (McDonald and Nivre, 2007). Lastly, as mentioned above, *and Mary coffee* forms a syntactic unit, which is not captured by this proposal.

6.2 Gerdes and Kahane (2015)

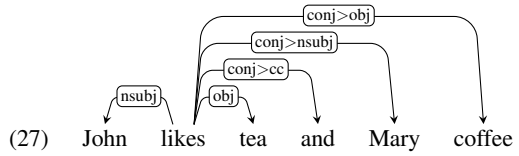
Gerdes and Kahane (2015) propose a graph-based analysis of gapping constructions, which inspired the analysis of gapping constructions in UD v2. Their proposal is, by and large, a combination of the "remnant" analysis and the "orphan" analysis that we described in this paper. They further add a `lat-NCC` (lateral non-constituent coordination) relation between the correspondents in the clause with the overt verb. According to their proposal, we would analyze a sentence with a single verb gap as follows.[6]



The advantage of this proposal is that it captures two different things: The `orphan` relation captures the fact that *and Mary coffee* forms a syntactic unit and the `remnant` relations allow one to determine the type of argument of each orphan. Nevertheless, this proposal also comes with several drawbacks. First, this analysis leads to graphs that are no longer trees and it is therefore not suited for the *basic* UD representation, which is supposed to be a strict surface syntax tree (Nivre et al., 2016). This would not be an issue for the *enhanced* representation, which may be a graph instead of a tree, but the `remnant` relations in this analysis can lead to the same problem as mentioned above. That is, if an orphan does not have a correspondent within the same sentence, we cannot use this type of dependency. Further, the copy nodes in our *enhanced* representation capture the fact that this sentence is describing two distinct "liking" events, which is not captured in this analysis. Finally, while their `lat-NCC` relation seems unproblematic from a theoretical point of view, we do not see its advantage in practice. For these reasons, we adopt only part of their proposal for the *basic* representation and introduce copy nodes in the *enhanced* representation.

6.3 Composite relations

Joakim Nivre and Daniel Zeman developed a third proposal[7] as part of the discussion of the second version of the UD guidelines. Their proposal is based on composite relations such as `conj>nsubj`, which indicates which relations would be present along the dependency path from the first conjunct to the orphan if there was no gap. For example, `X conj>nsubj Y` indicates that there would have been a `conj` relation between X and an elided node, and an `nsubj` relation between the elided node and Y. According to this proposal, we would analyze a sentence with a single verb gap as follows.



The advantages of this proposal are that the relation names provide much more information on the type of dependent than the generic `orphan` relation, and that in most cases, the *enhanced* representation can be deterministically obtained by

[6] We translated their relation names to the appropriate UD relations to make it easier to compare the various proposals.

[7] See http://universaldependencies.org/v2_prelim/ellipsis.html for a more detailed description of their proposal.

splitting up the relation name and inserting a copy of the governor of the composite relation. For example, for the above sentence, one could obtain the *enhanced* representation by copying *likes* and attaching *and*, *Mary*, and *coffee* with a *cc*, *nsubj*, *obj* relation, respectively.

However, this representation also comes with several drawbacks. First, it drastically increases the size of the relation domain as in theory, an unbounded number of relations can be concatenated. For example, in the sentence in (12d), we would end up with a `conj>xcomp>xcomp>xcomp>obj` relation[8] between *want* and *play*. This is highly problematic from a practical point of view as virtually all existing parsers assume that there is a finite set of relations than can appear between two words. Further, such an analysis also introduces many more long-distance dependencies. For these reasons, it seems unlikely that a parser would be able to produce this representation (and consequently also not the *enhanced* representation) with high accuracy. Finally, also in this case, the second conjunct does not form a syntactic unit.

To summarize this comparison, the main drawback of the *orphan* analysis is that it does not capture any information about the type of arguments in the basic representation. Despite this drawback, we believe that the analysis of gapping constructions in UD version 2 is better with regard to theoretical and practical considerations than any of the previous proposals, because (a) it can be used to analyze sentences in which orphans do not have correspondents; (b) it does not increase the number of relations; (c) it does not introduce additional long-distance dependencies or non-projective dependencies; and (d) it captures the fact that the second conjunct forms a syntactic unit.

[8] As pointed out by a reviewer, one could limit the relation name to the first and last relation in the hypothetical dependency path, e.g., `conj>obj` in this example. While this would put an upper bound on the number of relations, it would no longer be possible to deterministically obtain the *enhanced* representation in these cases. Further, we would assign the same relation label to different arguments in some cases. For example, we would add a `conj>obj` relation between *want* and *John* and between *want* and *play* in the analysis of the following sentence despite the fact that these two phrases are arguments to different verbs.

(28) Mary wants Tim to write a novel, and (wants) John (to write) a play.

7 Conclusion and future directions

We discussed which kind of gapping constructions are attested in a variety of languages, and we provided a detailed description how these constructions can be analyzed within the UD version 2 framework. We further explained how sentences with gaps can be analyzed in the *enhanced* UD representation, and we argued why we believe that the current proposal gives the best tradeoff between theoretical and practical considerations.

While we discussed what *enhanced* UD graphs of sentences with gapping should look like, we did not provide any methods of obtaining these graphs from sentences or basic UD trees. One future direction is therefore to develop methods to automatically obtain this representation.

Acknowledgments

We thank the entire UD community for developing the guidelines and for providing feedback to various proposals as part of the the discussion of the second version of the UD guidelines. We also thank Joakim Nivre for providing comments on an early draft of this paper. Further, we thank the anonymous reviewers for their thoughtful feedback. We especially appreciated the thorough and insightful feedback by the second reviewer, and while we were not able to incorporate all of it due to space and time constraints, we hope to address most of the remaining points in future versions of this work. This work was supported in part by gifts from Google, Inc. and IPSoft, Inc. The first author is also supported by a Goodan Family Graduate Fellowship.

References

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. 2016. Learning to compose neural networks for question answering. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2016)*.

Gabor Angeli, Melvin Johnson Premkumar, and Christopher D Manning. 2015. Leveraging linguistic structure for open domain information extraction. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP)*.

Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová,

Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, and Šárka Zikánová. 2013. Prague Dependency Treebank 3.0. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University.

Cem Bozsahin. 2000. Gapping and word order in Turkish. In *Proceedings of the 10th International Conference on Turkish Linguistics.*

Elizabeth Coppock. 2001. Gapping: In defense of deletion. In Mary Andronis, Christopher Ball, Heidi Elston, and Sylvain Neuvel, editors, *Papers from the 37th Meeting of the Chicago Linguistic Society*, Chicago. Chicago Linguistic Society.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. Stanford typed dependencies manual. Technical report, Stanford University.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Annahita Farudi. 2013. *Gapping in Farsi: A Crosslinguistic Investigation*. Ph.D. thesis, MIT.

Kim Gerdes and Sylvain Kahane. 2015. Non-constituent coordination and other coordinative constructions as dependency graphs. In *Proceedings of the Third International conference on Dependency Linguistics (Depling 2015)*.

Jane Grimshaw. 1997. Projection, heads, and optimality. *Linguistic Inquiry*, 28(3):373–422.

Ray S. Jackendoff. 1971. Gapping and related rules. *Linguistic Inquiry*, 2(1):21–35.

Kyle Johnson. 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry*, 40(2):289–328.

Kumiko Kato. 2006. *Japanese Gapping in Minimalist Syntax*. Ph.D. thesis, University of Washington.

Dave Kush. 2016. Notes on gapping in Hindi-Urdu: Conjunct size and focus parallelism. *Linguistic Analysis*, 40(3-4):255–296.

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Ryan McDonald and Joakim Nivre. 2007. Characterizing the errors of data-driven dependency parsing models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Igor A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. State University of New York Press, Albany.

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. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Timothy Osborne. 2006a. Gapping vs. non-gapping coordination. *Linguistische Berichte*, 207:307–337.

Timothy Osborne. 2006b. Shared material and grammar: Toward a dependency grammar theory of non-gapping coordination for english and german. *Zeitschrift für Sprachwissenschaft*, 25(1).

Timothy Osborne. 2015. Diagnostics for constituents: Dependency, constituency, and the status of function words. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*.

Carl Pollard and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press.

Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata. 2017. Universal semantic parsing. *arXiv preprint arXiv:1702.03196*.

John Robert Ross. 1970. Gapping and the order of constituents. In Manfred Bierwisch and Karl Erich Heidolph, editors, *Progress in Linguistics*. De Gruyter, The Hague.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning (CoNLL 2008)*.

Lucien Tesnière. 1959. *Eléments de Syntaxe Structurale*. Klincksieck, Paris.

G. Vanden Wyngaerd. 2007. Gapping constituents. Ms., FWO/K.U. Brussel.