

Coconstructions in Spoken Data: UD Annotation Guidelines and First Results

Ludovica Pannitto¹, Sylvain Kahane², Kaja Dobrovoljc³,
Elena Battaglia¹, Bruno Guillaume⁴, Caterina Mauri¹, Eleonora Zucchini⁵

¹ University of Bologna, Bologna - Italy;

² Université Paris Nanterre, Modyco, Paris - France;

³ University of Ljubljana and Jozef Stefan Institute, Ljubljana - Slovenia;

⁴ Université de Lorraine, CNRS, Inria, LORIA, Nancy - France;

⁵ Masaryk University, Brno - Czech Republic

Abstract

The paper proposes annotation guidelines for syntactic dependencies that span across speaker turns — including collaborative coconstructions proper, wh-question answers, and backchannels — in spoken language treebanks within the Universal Dependencies framework. Two representations are proposed: a speaker-based representation following the segmentation into speech turns, and a dependency-based representation with dependencies across speech turns. New propositions are also put forward to distinguish between reformulations and repairs, and to promote elements in unfinished phrases.

Keywords: spoken language, coconstruction, backchannel, reformulation, syntactic treebank, French, Italian, Slovenian

1. Introduction

Spoken language is intrinsically temporal, incremental, and interactive (e.g., Hopper 1992; Auer 2024; Deppermann and Günthner 2015): unlike written text, it unfolds in real time and is jointly produced by multiple participants cooperating in the construction of syntactic structure. Linearized transcriptions, however, necessarily flatten this temporal and multi-party dimension.

Universal Dependencies (UD, de Marneffe et al. 2021), despite its success as a cross-linguistic annotation framework, inherits a broadly shared assumption in syntactic theory: that syntax operates within the boundaries of a sentence. While this assumption is largely unproblematic for written language (although sentence boundaries detection is not yet considered a completely solved task, e.g. Rudrapal et al. 2015; Read et al. 2012; Sprugnoli and Redaelli 2025), it raises difficulties for spoken interaction, where segmentation into sentence-like units is neither trivial nor theoretically neutral.

Two additional issues complicate the picture. First, segmenting the speech flow is not equivalent to segmenting speaker turns: different segmentation strategies may be adopted depending on whether the focus is on individual speakers' production or on jointly constructed syntactic units. Second, syntax can be viewed both as a cognitive mechanism for building grammatical structures and as an interactional resource for coordinating meaning between speakers. These perspectives stem from different research traditions, yet an an-

notation framework such as UD should not commit exclusively to either one.

We introduce a novel proposal for annotation guidelines that can enable resource creators and users to account for syntactic dependencies holding among speech produced by different speakers. In Section 2, we argue that spoken data require an explicit treatment of syntactic relations that cross turn boundaries. We refer to these relations as *coconstructions*. We provide a typology of such constructions in Section 3, and we propose annotation guidelines that preserve information from both speaker-based and dependency-based perspectives, while remaining fully compatible with the UD formalism, in Section 4. Section 5 presents the conversion between the speaker-based and the dependency-based representation. A quick overview of the implementation of our annotation scheme in three treebanks from French, Italian, and Slovenian is provided in Section 6.

2. The Units of Spoken UD Treebanks

No clear agreement emerges in the literature about what a *sentence* (as in, maximal unit of segmentation) in spoken language is (Sacks et al., 1974; Pietrandrea et al., 2014; Mettouchi and Vanhove, 2020) and, despite sentences being the main building blocks of UD resources, no explicit definition of what should count as a sentence is ever given in the current annotation guidelines.

Syntactic annotation frameworks typically operate on segmentation units that resemble sentences but can vary greatly in their definitions. This mis-

match becomes particularly evident when syntactic dependencies span across speaker turns, overlap with other speakers' contributions, or are resumed after interruptions. Pietrandrea et al. (2014) recall three commonly applied criteria employed for the definition of sentences, namely the locutionary criterion, that attributes sentences to the locutionary responsibility of a given speaker; the graphic or prosodic criterion, that defines sentences based on punctuation markers or major prosodic breaks; and the syntactic criterion, that focuses on sentences as maximal structurally autonomous units, linked to the surrounding context by discursive relations.

While for standard written data the three criteria are likely to converge to the same analysis, this is not the case for spoken language and other varieties such as user-generated content (Sanguinetti et al., 2023). Earlier attempts at building spoken treebanks annotated for constituent syntax have implemented a turn-based approach. For instance, in the Penn Treebank-style annotation of the Switchboard corpus, turns by different speakers are parsed as standalone trees (Bies et al., 1995); in the extended Switchboard Dialog Act Corpus - SwDA (Jurafsky et al., 1997), by representing a discourse layer on top of a syntactic one, only some forms of co-construction, for instance collaborative completions, receive a tag. Another example is the Verbmobil corpus, where the primary domain of syntactic analysis and annotation is “the dialog turn, (...) an uninterrupted contribution by one dialog participant” (Hinrichs et al., 2000, 108); while this approach can account for some phenomena of spoken language such as disfluencies (see Meteer et al. 1995), it privileges single-speaker syntax and obscures collaborative syntax, which cannot be properly searched within trees. Dependency frameworks are, as a matter of fact, more apt to represent relations holding across turn boundaries. However, among existing spoken language treebanks in UD, no uniform segmentation criterion is applied, probably as a result of pre-existing design choices in the source data (Dobrovoljc, 2022).

We argue that two different units of analysis need to be considered in the case of spoken data:

- **speaker-dependent maximal units**, loosely related to the notion of *turn* and that anchor segmentation to the speech production of each speaker participating in the conversation, as it is usually done;
- **structurally-autonomous maximal units**, based on the notion of government (Tesnière, 1959, 2015; Blanche-Benveniste, 1987), defined by Kahane and Pietrandrea (2012) as the property by which words impose constraints on other words, specifically constraints on (i) their syntactic category, (ii) their morphologi-

cal or syntactic marking, and (iii) their linear position.

An alternative segmentation strategy, adopted for instance in Rhapsodie, is based on the notion of illocutionary unit, defined through prosodic and pragmatic criteria. Such units capture the organization of speech in terms of illocutionary completeness and prosodic packaging, and therefore reflect an important dimension of spoken interaction. However, because they are grounded in prosodic-pragmatic segmentation, they do not necessarily align with domains of syntactic dependency. As a result, using illocutionary units as primary units for syntactic annotation would make the representation of syntax dependent on a specific analysis of prosodic and illocutionary organization, which is not always available, consistently applicable, or comparable across corpora and languages. By contrast, speaker-dependent maximal units provide a more neutral and operational point of departure: they are directly anchored in the observable distribution of speech across participants, preserve information about production and turn organization, and can be systematically enriched to recover syntactic dependencies that extend across unit boundaries, without conflating speaker organization, prosodic packaging, and syntactic structure into a single segmentation criterion.

In the remainder of this paper, we refer to speaker-dependent maximal units as *maximal units* and to structurally-autonomous maximal units as *rectional units*. Currently, spoken treebanks in UD are organized in speaker-dependent units, and only Rhapsodie (see Section 6.1) and KIParla Forest (see Section 6.2) provide special features for cases of coconstructions.

3. A Typology of Coconstructions

We consider as *coconstructions* those syntactic relations (without restrictions on dependency type) that hold between utterances produced (i) by different speakers or (ii) by the same speaker after an interruption by another participant. In interactive scenarios, speakers may jointly contribute to the realization of a single syntactic structure, sometimes in a tightly coordinated fashion and sometimes with partial or even complete overlap (Lerner, 1991; Ono and Thompson, 1996; Sacks, 1992; Helasvuoto, 2004; Calabria, 2026). From a dependency-based perspective, this implies that syntactic relations cannot always be confined to a single speaker turn.

Our notion of coconstruction is grounded in a dependency-oriented view of syntax and is compatible with the concept of rectional unit (*unité rectionnelle*) introduced by Benzitoun et al. (2010); Kahane and Pietrandrea (2012), described in Section 2: this is defined independently of speaker

turns or prosodic segmentation and it captures syntactic governance relations wherever they occur in the speech flow. This makes the notion particularly well suited to spoken interaction, where the realization of a single rectional domain may be distributed across multiple speakers or interrupted by other contributions. Coconstructions can thus be understood as cases in which a single rectional unit is jointly realized across turns, by one or more speakers.

On this basis, we distinguish three major types of coconstructions, reflecting different interactional and syntactic configurations: *co-construction proper*, *wh-question constructions*, and *backchanneling*.

3.1. Coconstruction Proper

Coconstruction proper covers cases in which multiple contributions target the same dependency structure by completing, duplicating, or extending dependencies. Such contributions can be produced (i) by different speakers or (ii) by the same speaker after an interruption by another speaker, i.e. in non-contiguous utterances, which formally results in the same pattern as (i)¹. In the remainder of the article we will refer to Speaker 1 and Speaker 2 as the producers of the first and second unit which are focus of the discussion. Here and after, in the event of an interruption, Speaker 2 coincides with Speaker 1. All examples are reproduced preserving the original transcription conventions of the corpus.

The key distinction within coconstructions proper concerns whether Speaker 2 realizes a dependency that is already projected by Speaker 1 (case A), or whether Speaker 2 adds a new dependency to the structure initiated by Speaker 1 (case B).

A. Speaker 2 realizes an already projected dependency. When Speaker 2 realizes a dependency projected by Speaker 1, we distinguish three subcases reflecting the status of that dependency at the moment of realization: unrealized, in progress, or already realized.

A.1 Completion: Speaker 1 leaves a syntactic dependency unrealized, and Speaker 2 provides the missing material, *completing* the structure initiated by Speaker 1, as in Example (1);

- (1) English (Ono and Thompson 1996, p. 72)
 L: his position is **pretty uh**
 A: ... **stable**.

¹We do not treat the case of contiguous maximal units produced by the same speaker, as we hypothesize that these should be merged in a single unit in case they can be linked through syntactic dependency, thus yielding a single maximal rectional unit.

A.2 Parallel realization: The same syntactic dependency is realized at the same time by both speakers. Speaker 1 projects and eventually completes the structure, often after hesitation or a false start, while Speaker 2 simultaneously or near-simultaneously fills the same syntactic slot, as in Example (2), see also Figure 3 in Appendix A. This configuration frequently results in overlap;

- (2) Italian (KIParla², BOD2018)
 BO140: mh: vabbè [ci **stiamo**] **insie:me:**
well we stay together there
 esatto: c'è il divano comodo,
right there's a comfortable sofa
 BO118: [**state insie:me:**]
stay together

A.3 Subsequent realization: Speaker 1 completes their own syntactic structure, and Speaker 2 subsequently selects a dependency in Speaker 1's structure and realizes it again, either by repeating or replacing material in the same syntactic slot (as in Examples (3) and (4)).

- (3) French (Rhapsodie, D2001)³
 \$L2 "eh bien" je crois que je ne me suis pas
well I think I haven't
 conduit d'une façon conforme à ce qu'on
behaved as one would
 attend "euh" { { **d'une jeune fille d'abord**
expect from a young girl first
 | ^ **et d'une femme ensuite** } } //+
and from a woman afterwards
 \$L1 { | **d'une jeune bourgeoise** } ?//+
from a young bourgeois girl?
 \$L2 { | "**disons**" **d'une jeune bourgeoise** } //
let's say from a young bourgeois girl
- (4) Italian (KIParla, KPS021)
 PKP126: quindi l- la linea **tra finzione e realtà**
so the line between fiction and reality
 cioè tra verità non verità
that is between true and false
 non ho ancora capito [dove sta]
I still haven't understood where it lies
 PKP125: [**più tra de]tto e non detto**
more between said and unsaid

²(Mauri et al., 2019). In all examples from KIParla, square brackets indicate overlaps between speakers' productions.

³Curly brackets indicate conjuncts in a stacking construction. See Section 4.2. The sign // indicates the end of an illocutionary unit. The sign //+ indicated that there is a link to the next unit.

B. Speaker 2 adds a new dependency. Speaker 2 realizes a syntactic dependency that is *not* already projected by Speaker 1, thereby *extending* the syntactic structure initiated by Speaker 1 (Examples (5), (6), (9))

(5) English (Ono and Thompson 1996, p. 81)

M: they must know each other.

H: ... **very well.**

(6) Italian (KIParla, BOA3017)

BO139: quando si parlano sopra
when they speak one over the other
devi mettere delle parentesi quadre
you have to put square parentheses
intorno **alle parole**
around words

BO146: ah okay
oh okay

BO139: **che si sovrappongono**
that overlap

BO136: **come nei sottotitoli**
like in subtitles

3.2. Wh-questions

Answers to *wh*-questions are analysed as syntactically dependent on the question, and thus as part of the same rectional unit, as already proposed in the guidelines of Rhapsodie.

(7) Italian (KIParla, PTA007)

TOR001: **dove** vai ad arrampicare?
where do you climb?

TOI007: **al bi side**⁴ vicino alla colletta
at the bi side close to colletta

We therefore consider answers to *wh*-questions as cases of coconstructions, insofar as they realize syntactic material licensed by the interrogative structure of the preceding turn (see an example in Figure 6). By contrast, we do not treat answers to polar questions as coconstructions, since their response items do not depend syntactically on any element of the question nor fill a pre-existing syntactic slot (ex., cf. Example (7) with A: *do you climb at the bi side?* B: *yes*).

3.3. Backchannelling

Backchannels are short productions uttered by one participant while another speaker holds the floor. They may be verbal or paraverbal (e.g. *mhm* as in Example (8), *yes* as in Example (9), *right*) and

⁴'bi side' is the quasi-phonetic rendering of the name of a climbing gym, 'Bside'.

mainly serve to signal attention or alignment (Ward and Tsukahara, 2000; Mereu et al., 2024; Furkó, 2020; Khan, 2023). At their core, these verbal reactions show that the speaker has heard the contribution of the partner, often adding that it has been understood and accepted.

To be recognized as backchannels, such utterances (*i*) must be addressed to the content of the other speaker's contribution; (*ii*) must not be required or expected by the preceding turn (e.g. answers to *wh*- and polar questions are expected and required, so they cannot be considered backchannels, see (7)); (*iii*) must not require a reaction from the main speaker.

(8) Italian (KIParla, BOA3017)

BO145: ma perché mamma c'ha dei
but because my mom has
pregiudizi nei miei confronti
prejudices against me

BO139: **mhmh**
mhmh

BO145: e poi daniela non devi avere
and then daniela you shouldn't have
pregiudizi su di me
prejudice against me

(9) Italian, (KIParla, BOD2018)

BO118: sì sì ma anch'io però era proprio
yeah yeah same here but I really just
l'esigenza di stare da sola
needed to be alone

BO140: **sì** di stare da sola di non
yeah to be alone not
parlare
to talk

Backchannels differ from co-construction proper and from *wh*-questions coconstructions, in that they do not contribute syntactic material to the rectional unit projected by the main speaker. Nevertheless, they are tightly integrated into the temporal unfolding of speech and must be accounted for in the syntactic analysis of spoken interaction. Jucker and Smith (1996) argue that the distinction between a "discourse marker" and a "backchannel" is often more about conversational role than linguistic function and both belong to a broader category of elements used to negotiate "common ground".

We are aware that the syntactic status of backchannels is not uncontroversial. What counts as "syntax", however, is itself a matter of theoretical tradition: different frameworks draw the boundary between syntax, pragmatics, and interaction differently, and UD itself already encodes relations — such as *discourse* — that would not be considered syntactic under a narrow structural definition. More importantly, even if a backchannel is not itself

a syntactic dependent, it actively shapes the syntactic structure being built by the main speaker: it may prompt continuation, signal that repair or elaboration is needed, or mark the boundaries within which the syntactic unit unfolds. Ignoring backchannels in the annotation of spoken interaction thus risks misrepresenting the conditions under which syntactic structure is produced and interpreted. Including them in the `MISC` field — without altering any dependency relation — is a minimal and non-disruptive way of making this interactional layer visible to downstream users of the treebank.

4. Proposed Guidelines

To explicitly represent coconstructed configurations in UD annotation, we introduce two features encoded in the `MISC` field: `Coconstruct` and `Backchannel`. These features do not replace dependency relations, but provide a lightweight mechanism to signal linkage between speaker-based maximal units involved in coconstruction phenomena.

All features introduced in this paper are encoded in the `MISC` field, keeping the proposal fully backward-compatible: the core dependency structure, POS tags, morphological features, and tokenization remain untouched. The `MISC` field is already used across treebanks for typographic information (`SpaceAfter=No`), transliteration, named entity spans, and discourse-level features; our proposal follows this established practice.

As a general principle, both `Coconstruct` and `Backchannel` are always assigned to the element that comes second in time: the feature is assigned to the responding token and its value identifies the eliciting or projecting token in the earlier utterance. This also enables multiple later tokens to point to the same earlier anchor.

Formally, `Coconstruct` encodes a backward pointer of the form `<deprel>::<sent_id>::<tok_id>`, while `Backchannel`, which already uniquely identifies a dependency relation (i.e., `discourse:backchannel`), encodes a backward pointer of the form `<sent_id>::<tok_id>` (see Figure 1).

The `Coconstruct` feature is used in two main configurations, which differ in the syntactic status of the dependency slot targeted by the responding token:

Completion of an open dependency slot: the responding token realizes a dependency that was projected but not realized in the preceding unit.

Stacking on an already realized dependency slot: the responding token provides an alternative

realization of a dependency that is already present in the preceding unit.

4.1. Completion of an Open Dependency Slot

A configuration qualifies as completion of an open dependency if all of the following conditions are met:

- Speaker 1 produces a token that syntactically licenses a dependent (although not necessarily requiring it), but the dependent is not realized before Speaker 2’s intervention.
- Speaker 1 does not complete the same dependency slot before or concurrently with Speaker 2’s contribution.

When these requirements are met, the head of the coconstructed dependency is the token in Speaker 1’s utterance that projects the unrealized dependency. Head identification follows standard UD principles: the head is the lexical item that selects the dependent.

This configuration applies both to cases A.1 and B described in Section 3. In cases of A.1 Speaker 1 might produce an element that projects a dependency but remains syntactically unfinished, because its canonical dependent is not realized before turn transfer. This situation is particularly frequent with adpositions (more specifically, prepositions), complementizers, and similar selecting elements. From a UD perspective, such cases raise a problem: the projected dependent is missing, but dependency annotation cannot be postponed or left underspecified. To address this, we adopt the promotion mechanism used in UD for ellipsis. When a projected dependency remains unrealized in Speaker 1’s utterance, the unfinished element is promoted to fill the syntactic position that would normally be occupied by its missing complement. This allows the dependency structure to remain well-formed while preserving information about the element’s incomplete status.

We introduce two additional `MISC` features to encode this mechanism: `Scrap=Yes` to mark any element (or subtree) whose syntactic projection remains unfinished and `Promotion=<deprel>` to indicate the dependency relation the element would have had before promotion, i.e. the relation it would bear to its governor if it were present.

- (10) Italian, (KIParla, BOA3017)
 BO146: ((ride)) vabbè te ti ci ti ci devo **fare**
 ((laughs)) *well you I have to do you*
pe::r
by
- BO139: ((ride)) **mestiere**
 ((laughs)) *trade*

For instance, in Example (10), a noun is missing after the preposition *per*. The oblique prepositional phrase is unfinished, so the *obl* dependency relation has to go on the preposition. The preposition, which should have been the *case* of the missing noun, has then been promoted. Since this promotion follows a *case* relation, the preposition receives the feature *Promotion=case*, as well as a feature *Scrap=Yes* indicating that the phrase it heads is unfinished (see first line in Figure 2).

Unfinished syntactic structures may arise in spoken interaction independently of whether co-construction is eventually realized. The use of *Scrap=Yes* therefore does not presuppose that the missing material will be supplied later: it encodes structural incompleteness as such, regardless of subsequent interactional developments. *Promotion* is essentially used in cooccurrence with *Scrap=Yes* when there is an unfinished phrase without its expected head.⁵ As we show in Section 5, the feature *Promotion* plays an important role in the conversion to the dependency-based representation.

The promotion of an element to a given position introduces an additional problem, because a dependent of a promoted word can be either a true dependent of the word itself or of its missing head. We mark this by means of a feature *Head*, with possible values *Word* and *Position*, to solve this problem. In Example (11), *nice* will be promoted as the *obj* of *has* and it receives a feature *Promotion=amod*, *a* is *det* of *nice* with a feature *Head=Position*, and *very* is *advmod* of *nice* with a feature *Head=Word*. In appendix, the full annotation of Example (11) is given (Figure 4) together with a Slovenian example (Figure 5) taken from the corpus.

- (11) A: she has a very nice
B: attitude

4.2. Stacking on an Already Realized Dependency Slot

A configuration qualifies as stacking on an already realized dependency slot if all of the following conditions are met:

- Speaker 1 realizes a complete dependency structure, including the dependent that fills the relevant syntactic slot;

⁵*Promotion* could, for instance, be introduced in case of gapping, as in (i) where *department* is promoted as root of the second conjunct.

(i) Mary called the police and Peter the fire department.

The feature *Promotion=obj* could be an elegant way to indicate that *department* is *obj* of an ellipsed verb.

- Speaker 2 produces material that is syntactically comparable to the dependent already realized by Speaker 1 and could occupy the same dependency position.

In stacking configurations, Speaker 2 does not complete an open dependency, but rather produces material that targets the same syntactic slot as an existing dependent. From a production perspective, Speaker 2's contribution is not structurally autonomous: its interpretation requires reference to Speaker 1's syntactic structure.

Stacking can however be instantiated by various *deprels*.

The most straightforward case is that of simple repetitions or repairs without any communicative intention, for which we recommend the *reparandum* relation. These are cases such as *des, des* or *ce dont, ce dont* in Example (13), or cases where a word is unfinished as *il sotto~* 'the sub~' in Example (12) (see also Figure 7): the first speaker has difficulties to produce the word *sottotitolatore* 'subtitled', which is then correctly produced by the second speaker. *reparandum* is a right-headed relation in UD: in order to mark a coconstruction on the latter item of the relation, we introduce the special *deprel repair* which is converted to *reparandum* in dependency-view.

- (12) Italian (KIParla, BOA3017)
BO147: infatti te potresti fare **il sotto~**
indeed, you could be the subti~
quello che fa i sottotitoli
the one that does subtitles
BO145: **sottotitolatore**
subtitled

If instead the first conjunct is complete and the second conjunct is different, we consider that we have a reformulation. Example (4) is a typical case: The first speaker already reformulates *tra finzione e realtà* 'between fiction and reality' by *tra verità non verità* 'between true and false' and then the second speaker proposes a second reformulation *più tra detto e non detto* 'more between said and unsaid'.

We propose to introduce a new relation for reformulation, which we call *conj:reform*. Contrary to coordinated conjuncts, where conjuncts have different referents, in a reformulation, the conjuncts are different denotations of the same referent (Kahane and Pietrandrea, 2012). We extend the relation *conj:reform* to other cases where two conjuncts target the same referent.

The first case is confirmation. In Example (3), speaker \$L2 produces *d'une jeune fille d'abord et d'une femme ensuite* 'from a young girl first and from a woman afterwards' and \$L1 proposes the reformulation *d'une jeune bourgeoise* 'from a young

bourgeois girl’ (with an interrogative prosody). Then \$L2 confirms this proposition by repeating it (with the addition of the discourse marker *disons* ‘let’s say’). The second case is the answer to a wh-question as in Example (7), where the interrogative pronoun *dove* ‘where’ is instantiated by the answer *al bi side* ‘at the bi side’. A third case is correction. We will illustrate it by the monologue in Example (13). In this Example, the speaker repeats three times *des Français* ‘of the French’, so there is no reformulation strictly speaking, but the second occurrence questions the choice of the proposed denotation, while the third, preceded by *pas seulement* ‘not only’, refutes it.

In all cases, $\langle \text{tok_id} \rangle$ points to the head of the first denotation.

- (13) French (Rhapsodie, D0004)
 \$L1: mais ce dont, ce dont vous parlez, c’est
but what, what you’re talking about is
 la crise générale des, des Français, oui,
the general crisis of the French, yes,
 enfin, des Français, pas simplement
well, of the French, not just
 des Français, hein, des, de l’humanité
of the French, but of the, of humanity
 et de la lecture.
and reading.

5. Automatic Conversion between Speaker-based and Dependency-based Representation

The two proposed annotation views (speaker-based and dependency-based) rely on different maximal segmentation principles. As UD and SUD annotation schemas strongly rely on the notion of ‘sentence’ for segmentation, we propose providing the two views separately. The speaker-based view, properly enriched with the features introduced in Section 4, can be converted into the dependency-based view, which relies on the definition of ‘sentence’ as ‘rectional unit’, enabling coconstructions to be realised as regular syntactic dependencies.

In the Figure 1, the first line is the speaker-based view and the last line is the dependency-based view. An intermediate richer structure is introduced in which: new dependency relations are identified with a suffix */attach* and they are drawn in green. The richer structure helps to understand the conversion process and adds the possibility to query these modified dependencies specifically. Note also that metadata relative to the speaker, which would usually be placed at sentence level, is now moved to tokens, to take into account the fact that a dependency-based sentence can include tokens uttered by different speakers.

As explained above, there may be other dependency relations needing to be restored in dependency-based view. The new features `Promotion` and `Head` are used to reconstruct the expected dependency structure. The Figure 2 shows an Italian example where the oblique prepositional phrase *per mestiere* is built by two different speakers. In this case, we mark, on *per* that the structure is unfinished (`Scrap=Yes`) and that it has been promoted (`Promotion=case`). The second line in the figure show the intermediate (graph) structure where both dependencies on *per* are drawn: in red with suffix `sb` for the promoted relation and in green with suffix `attach`, the expected UD relation dependency-based view. The last line gives the UD regular annotation of the dependency-based view. Similar examples are given in appendix for English (Figure 4) and for Slovenian (Figure 5).

Note that when the promoted token has its own dependencies, they may have to be lifted up in the final dependency-based structure. The figure 4 gives such an example on a built English example *this is pretty* in which *pretty* is a separate maximal unit in speaker-based view.

6. Coconstructions in Existing UD Treebanks

6.1. French (Rhapsodie)

Coconstructions and backchannels were already annotated in the first version of Rhapsodie (Benzitoun et al., 2010; Lacheret et al., 2014), using the delimiter `//+`, as in Example (3). When the treebank has been converted into UD, they were lost, and then reintroduced with a feature `AttachTo` for the address of the governor and `Rel` for the relation (Kahane et al., 2021). For backchannels, the value of `Rel` was `discourse`. Backchannels in Rhapsodie currently include *yes-no* answers to questions. Instead of `conj:reform` and `reparandum`, seven different labels were first considered (Kahane and Pietrandrea, 2012), but due to a quite poor inter-annotator agreement, they were merged into a unique relation, i.e., `conj:dicto`.

Among the 3,209 trees in the speaker-based version of Rhapsodie, 229 are annotated as backchannels, but there are also 93 interjections as roots that are potential backchannels and need to be revised. The treebanks also contain 35 coconstructions: 24 `conj:reform` and 11 modifiers, including two relative clauses.

6.2. Italian (KIParlaForest)

KIParla Forest (Pannitto et al., 2025) was first introduced in UD v2.17, based on the broader KIParla Corpus of Spoken Italian (Mauri et al., 2019; Ballarè

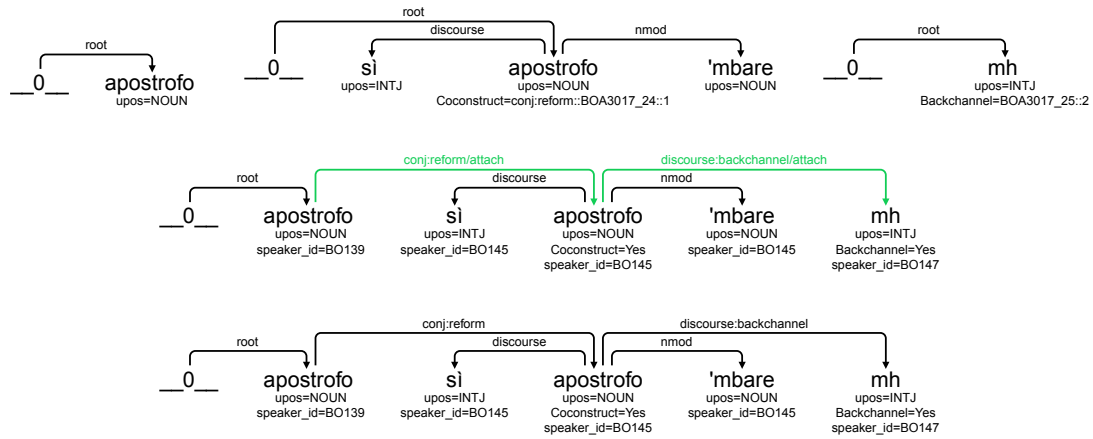


Figure 1: An Italian example of a `conj:reform` coconstruction and a backchannel. The example is taken from KIParla, BOA3017. Tr. en. ‘apostrophe’, ‘yes, apostrophe ‘mbare’, ‘mh’. ‘mbare is a dialectal form for *compare*, ‘mate’.

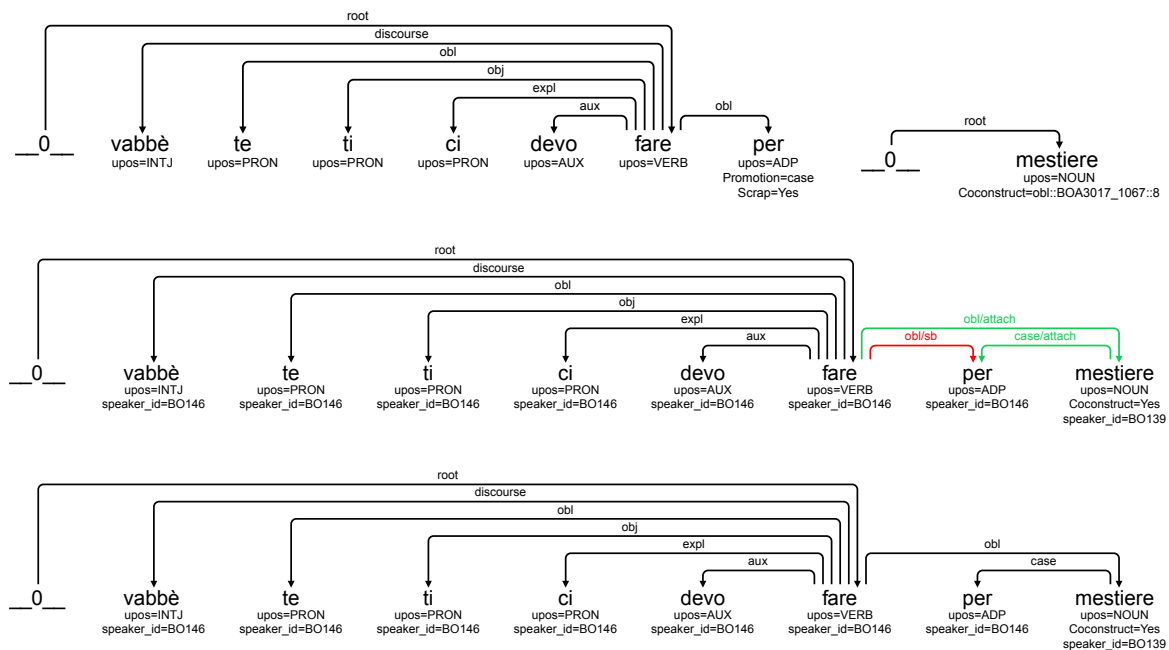


Figure 2: Italian example (10) with promotion (NB: the sentence is shortened, for space reasons). The first line shows speaker-based annotation, the second line shows the intermediate graph representation of the dependency-based view and line 3 shows final UD well-formed tree for the dependency-based view.

and Mauri, 2020).

The treebank currently hosts 2 conversations from the original corpus, an interview (BOD2018, comprising 2 speakers) and a free conversation (BOA3017, comprising 4 speakers), for a total of 9,135 tokens and 1,007 sentences.

Backchannels are annotated in 134 sentences: most of these are sentences composed by one single token that is frequently an INTJ such as *ah*, *eh*, *mh* or *ok* (67 cases), an ADV as *sì* ‘yes’, *no*, *certo* ‘certainly’, *ecco* ‘that’s it’, *esatto* ‘exactly’ (51 cases). Another significant group is that of

exact repetitions (4 cases) Also exclamations such as *porca troia* ‘holy shit’ are treated as cases of backchannels, and longer sequences such as *ah me l’hai raccontato* ‘right you told me’.

Currently the treebank contains 70 annotations of coconstructions. Most cases (38) concern conjuncts, either proper `conj` or `conj:reform` relations. `obl` and `acl:*` also constitute a big group (13 cases), as they capture the prototypical case of expansion of the dependency tree during interaction.

6.3. Slovenian (SST)

In the original compilation of the Spoken Slovenian Treebank (SST, Dobrovoljc and Nivre 2016; Dobrovoljc 2025), coconstructions were not treated explicitly. The treebank followed the speaker-based utterance segmentation of the GOS corpus (Verdonik et al., 2024), which encompasses 6,121 utterances and 98,393 tokens, and syntactic annotation was applied independently within each utterance. To evaluate the feasibility of the proposed guidelines and estimate the extent of coconstructions in Slovenian spoken data, we therefore began by targeting two configurations: completion of an unrealized dependency (Type A.1 in Section 3.1) and backchannelling (Section 3.3).

Backchannels were identified semi-automatically by extracting consecutive utterance pairs by different speakers where the first was not interrogative and the second contained one or more backchannel candidates (based on a preexisting list derived from items already marked as `discourse`, `INTJ` and/or `PART`). This procedure yielded 396 cases of backchannelling (69 different types), mostly single words (e.g. *ja* 'yes', *mhm* 'mm-hm', *aha* 'I see', *aja* 'oh, right', *dobro* 'okay'), but also longer sequences (e.g. *tako, tako, tako, tako* 'right, right, right, right').

For coconstruction proper (type A.1), we manually inspected utterance pairs where the first did not end in final punctuation, indicating potential incompleteness, and manually identified 17 cases of completed projected dependencies, including argument insertion (Speaker 1: *če atom odda elektrone, nastane...* 'if an atom gives away electrons, there arises...' – Speaker 2: *...kation* '...a cation') and sentence-level completion (Speaker 1: *kdor ti bo reklamo poslal...* 'whoever sends you an advertisement...' – Speaker 2: *...zaračuna* '...charges money').

These results provide an initial empirical basis for studying coconstructions in Slovenian and a retrieval procedure that can be refined to capture additional types, or extended to other UD treebanks where such phenomena have not yet been explicitly annotated.

7. Conclusion

UD has been first developed on the basis of written corpora with the aim of developing parsers (McDonald et al., 2013). The collection now includes spoken data for languages with a written tradition (as the three languages considered in this paper), as well as languages with only a spoken tradition (including sign languages). Due to the lack of clear guidelines for spoken data, important discrepancies exist among treebanks for spoken data (Dobrovoljc, 2022). This work is part of a common effort to propose guidelines and unify the annotation of

different treebanks. The first step in the syntactic analysis of spoken data is segmentation into minimal units (words or morphemes) and maximal units. The latter question is the main topic of this paper and the question of coconstructions is one of the very first problems to be solved. We hope to have convinced the reader of the interest in the question, as well as the relevancy of our propositions.

While the framework has been developed with spoken interaction in mind, its underlying mechanisms are more broadly applicable. Two extensions are worth noting. First, the proposed annotation scheme is potentially relevant for computer-mediated communication (CMC) data — such as instant messaging, or platforms like Slack or Discord — where users frequently fragment a single syntactic unit across multiple messages by sending partial utterances. These *split utterances* are structurally analogous to unfinished phrases in speech, and the `Coconstruct`, `Scrap`, and `Promotion` features could be applied to capture cross-message syntactic dependencies. A distinctive challenge of CMC, however, is the presence of *lag*: responses may not appear chronologically adjacent to the message they depend on, due to interleaved threads or parallel conversations. The pointer-based design of our features (encoding the identity of the eliciting token rather than relying on adjacency) makes the framework robust to such non-linear environments. Second, even within spoken data, the mechanism may be useful for cases of same-speaker continuation across deliberately maintained turn boundaries — for instance, when a speaker resumes a syntactic structure after a long backchannel sequence and the annotator wishes to preserve segmentation boundaries for prosodic or discourse reasons. In both cases, the features introduced here provide a lightweight but expressive mechanism for encoding syntactic continuity without collapsing the original segmentation.

We submit this proposal to the UD community hoping that it will be adopted as a standard extension for spoken and interactive treebanks, contributing to a more linguistically adequate treatment of multi-party and mediated language data.

Acknowledgements

This work received support from the CA21167 COST action UniDive, funded by the European Union via the COST (European Cooperation in Science and Technology), and from the ARIS research programme P6-0411.

References

- Peter Auer. 2024. [Online-Syntax](#). In Peter Auer, editor, *Online-Syntax: Eine Einführung in die Analyse gesprochener Sprache*, pages 1–19. Springer, Berlin, Heidelberg.
- Silvia Ballarè and Caterina Mauri. 2020. La creazione del corpus kiplarla: criteri metodologici e prospettive future. *Rivista Italiana di Dialettologia*, 44:53–69.
- Christophe Benzitoun, Anne Dister, Kim Gerdes, Sylvain Kahane, Paola Pietrandrea, and Frédéric Sabio. 2010. tu veux couper là faut dire pourquoi-propositions pour une segmentation syntaxique du français parlé. In *Congrès Mondial de linguistique française (CMLF)*, pages 2075–2090.
- Ann Bies, Mark Ferguson, Karen Katz, Robert MacIntyre, Victoria Tredennick, Grace Kim, Mary Ann Marcinkiewicz, and Britta Schasberger. 1995. Bracketing guidelines for treebank ii style penn treebank project. Technical report, University of Pennsylvania.
- Claire Blanche-Benveniste. 1987. *Pronom et syntaxe: l'approche pronominale et son application au français*. Peeters Publishers.
- Virginia Calabria. 2026. [The Grammar of Collaborative Turns: Temporality, Projection, Emergence and Clause Combination in Italian Talk-in-Interaction](#). De Gruyter Mouton.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. [Universal Dependencies](#). *Computational Linguistics*, 47(2):255–308.
- Arnulf Deppermann and Susanne Günthner. 2015. *Temporality in interaction*, volume 27. John Benjamins Publishing Company.
- Kaja Dobrovoljc. 2022. [Spoken Language Treebanks in Universal Dependencies: an Overview](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1798–1806, Marseille, France. European Language Resources Association.
- Kaja Dobrovoljc. 2025. Treebanking spoken slovenian: New data, models, and lessons learned. *Prispev. Novejso Zgod.*, 65(3).
- Kaja Dobrovoljc and Joakim Nivre. 2016. [The Universal Dependencies treebank of spoken Slovenian](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1566–1573, Portorož, Slovenia. European Language Resources Association (ELRA).
- Péter B. Furkó. 2020. *Discourse Markers and beyond: Descriptive and Critical Perspectives on Discourse-Pragmatic Devices across Genres and Languages*. Palgrave Macmillan.
- Marja-Liisa Helasvuo. 2004. Shared syntax: the grammar of co-constructions. *Journal of Pragmatics*, 36(8):1315–1336.
- Erhard W. Hinrichs, Julia Bartels, Yasuhiro Kawata, Valia Kordoni, and Heike Telljohann. 2000. The verbmobil treebanks. In *KONVENS 2000 / Sprachkommunikation, Vorträge Der Gemeinsamen Veranstaltung 5. Konferenz Zur Verarbeitung Natürlicher Sprache (KONVENS)*, 6. *ITG-Fachtagung "Sprachkommunikation"*, page 107–112, DEU. VDE-Verlag GmbH.
- Paul J Hopper. 1992. Times of the sign: Discourse, temporality and recent linguistics. *Time & Society*, 1(2):223–238.
- Andreas H. Jucker and Sara W. Smith. 1996. Explicit and implicit ways of enhancing common ground in conversations. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPRA)*, 6(1):1–18.
- Daniel Jurafsky, Elizabeth Shriberg, and Debra Bisasca. 1997. Switchboard swbd-damsl shallow discourse-function annotation: Coders manual. Technical report, Institute of Cognitive Science, University of Colorado, Boulder, CO, USA.
- Sylvain Kahane, Bernard Caron, Emmett Strickland, and Kim Gerdes. 2021. [Annotation guidelines of UD and SUD treebanks for spoken corpora: A proposal](#). In *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*, pages 35–47, Sofia, Bulgaria. Association for Computational Linguistics.
- Sylvain Kahane and Paola Pietrandrea. 2012. [La typologie des entassements en français](#). *SHS Web of Conferences*, 1:1809–1828.
- Shah Zeb Khan. 2023. The role of discourse markers, deixis, hedges, and backchannels in enhancing communication and emotional support. *Review of Law and Social Sciences*, 1(3):31–40.
- Anne Lacheret, Sylvain Kahane, Julie Beliao, Anne Dister, Kim Gerdes, Jean-Philippe Goldman, Nicolas Obin, Paola Pietrandrea, and Atanas Tchobanov. 2014. Rhapsodie: A prosodic-syntactic treebank for spoken french. In *Language Resources and Evaluation Conference (LREC)*.
- Gene H. Lerner. 1991. [On the syntax of sentences-in-progress](#). *Language In Society*, 20(3):441–458.

- Caterina Mauri, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti, and Francesco Suriano. 2019. KIParla Corpus: A New Resource for Spoken Italian. In *Proceedings of the Sixth Italian Conference on Computational Linguistics*.
- Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, et al. 2013. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97.
- Daniela Mereu, Francesco Cangemi, and Martine Grice. 2024. Backchannels are not always very short utterances. the case of italian multi-unit backchannels. *Journal of Pragmatics*, 228:1–16.
- Marie Meteer, Ann Taylor, Robert MacIntyre, and Ruth Iver. 1995. Disfluency annotation stylebook for the switchboard corpus. Technical report, Linguistic Data Consortium.
- Amina Mettouchi and Martine Vanhove. 2020. Prosodic segmentation and grammatical analysis in cross-linguistic corpora. In *42nd Annual Conference of the German Linguistic Society (DGfS), Workshop Corpus-based typology: spoken language from a cross-linguistic perspective*.
- Tsuyoshi Ono and Sandra A. Thompson. 1996. Interaction and syntax in the structure of conversational discourse: collaboration, overlap, and syntactic dissociation. In *Computational and conversational discourse*, pages 67–96, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Ludovica Pannitto, Eleonora Zucchini, Silvia Ballarè, Cristina Bosco, Caterina Mauri, and Manuela Sanguinetti. 2025. [Introducing KIParla forest: seeds for a UD annotation of interactional syntax](#). In *Proceedings of the Eighth International Conference on Dependency Linguistics (depling, Syntaxfest 2025)*, pages 54–73, Ljubljana, Slovenia. Association for Computational Linguistics.
- Paola Pietrandrea, Sylvain Kahane, Anne Lacheret-Dujour, and Frédéric Sabio. 2014. [The notion of sentence and other discourse units in corpus annotation](#). In Tommaso Raso and Heliana Mello, editors, *Spoken Corpora and Linguistic Studies*, pages 331–364.
- Jonathon Read, Rebecca Dridan, Stephan Oepen, and Lars Jørgen Solberg. 2012. [Sentence boundary detection: a long solved problem?](#) In *Proceedings of COLING 2012: Posters*, pages 985–994, Mumbai, India. The COLING 2012 Organizing Committee.
- Dwijen Rudrapal, Anupam Jamatia, Kunal Chakma, Amitava Das, and Björn Gambäck. 2015. [Sentence boundary detection for social media text](#). In *Proceedings of the 12th International Conference on Natural Language Processing*, pages 254–260, Trivandrum, India. NLP Association of India.
- Harvey Sacks. 1992. *Lectures on Conversation*. Blackwell, Oxford.
- Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.
- Manuela Sanguinetti, Cristina Bosco, Lauren Cassidy, Özlem Çetinoğlu, Alessandra Teresa Cignarella, Teresa Lynn, Ines Rehbein, Josef Ruppenhofer, Djamé Seddah, and Amir Zeldes. 2023. [Treebanking user-generated content: a UD based overview of guidelines, corpora and unified recommendations](#). *Language Resources and Evaluation*, 57(2):493–544.
- Rachele Sprugnoli and Arianna Redaelli. 2025. [Annotating manzoni: Challenges in the annotation of lemmas, POS and features in “I promessi sposi”](#). In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, pages 1084–1092, Cagliari, Italy. CEUR Workshop Proceedings.
- Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Klincksieck.
- Lucien Tesnière. 2015. *Elements of structural syntax*. John Benjamins Publishing Company.
- Darinka Verdonik, Kaja Dobrovoljc, Tomaž Erjavec, and Nikola Ljubešić. 2024. [Gos 2: A new reference corpus of spoken Slovenian](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7825–7830, Torino, Italia. ELRA and ICCL.
- Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in english and japanese. *Journal of pragmatics*, 32(8):1177–1207.

A. Annotation Examples

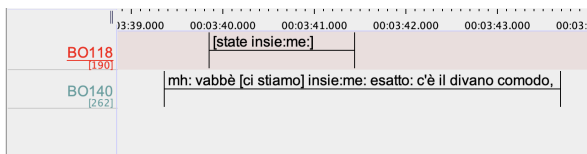


Figure 3: Original transcription of Example (10) in the KIParla corpus.

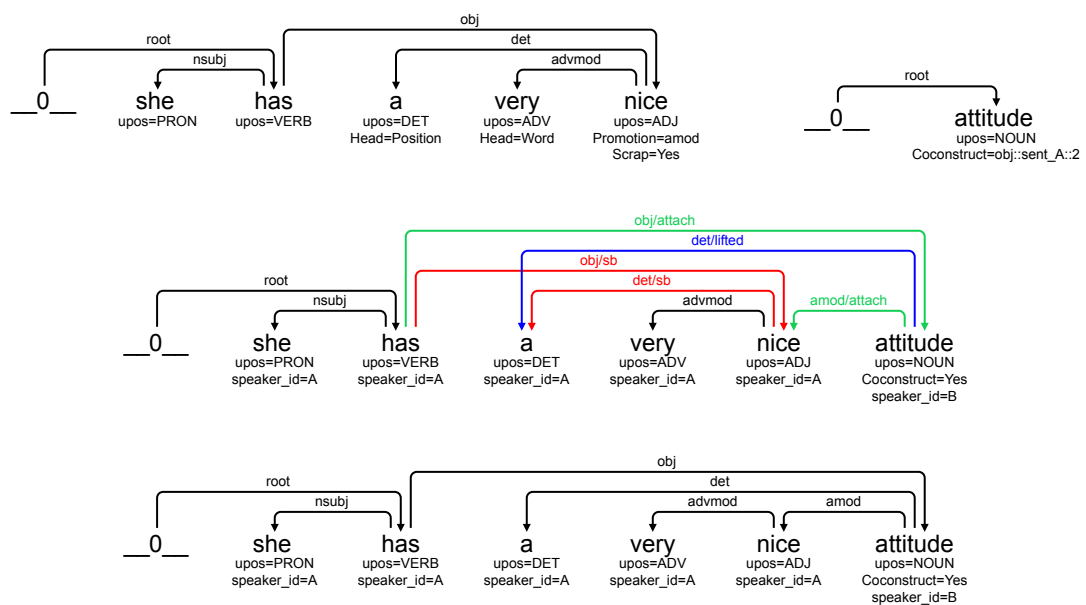


Figure 4: English example with lifting.
 Line 1: speaker-based annotation;
 Line 2: intermediate representation of the dependency-based view;
 Line 3: final UD tree for the dependency-based view

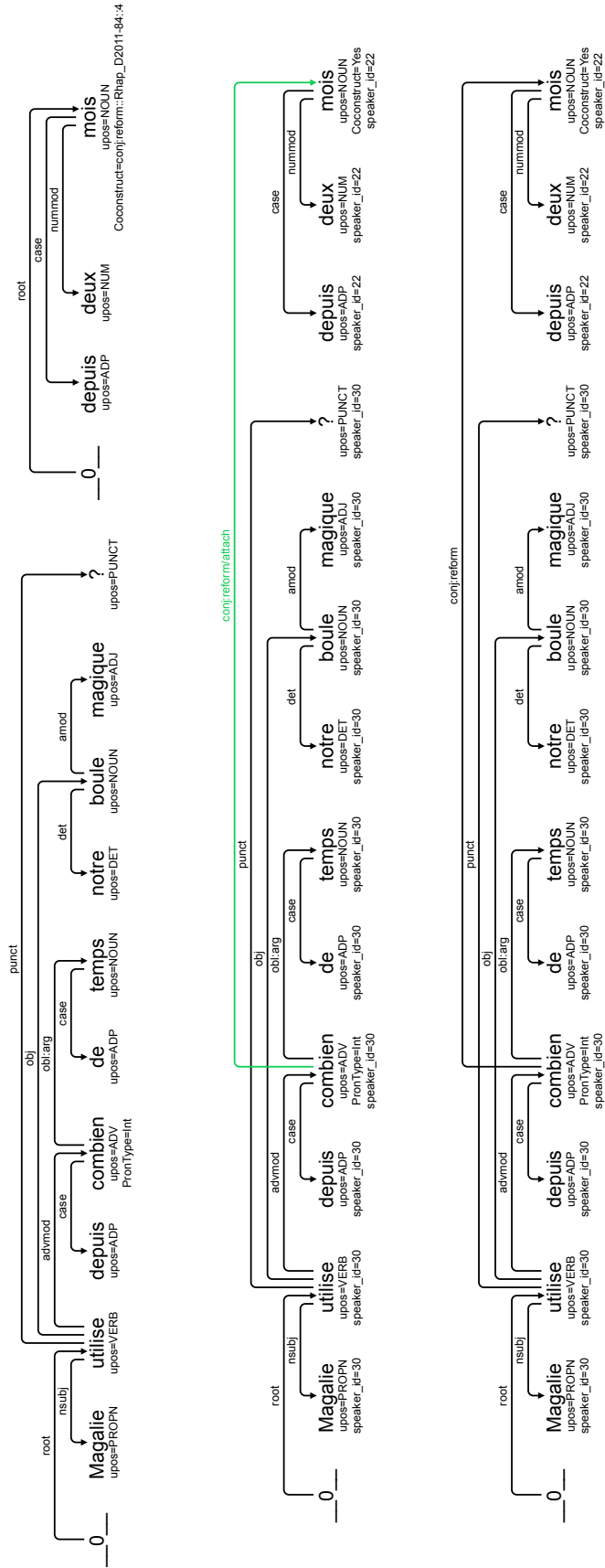


Figure 6: French example with a question
 'How long has Magalie been using our magic ball?', 'for two months'

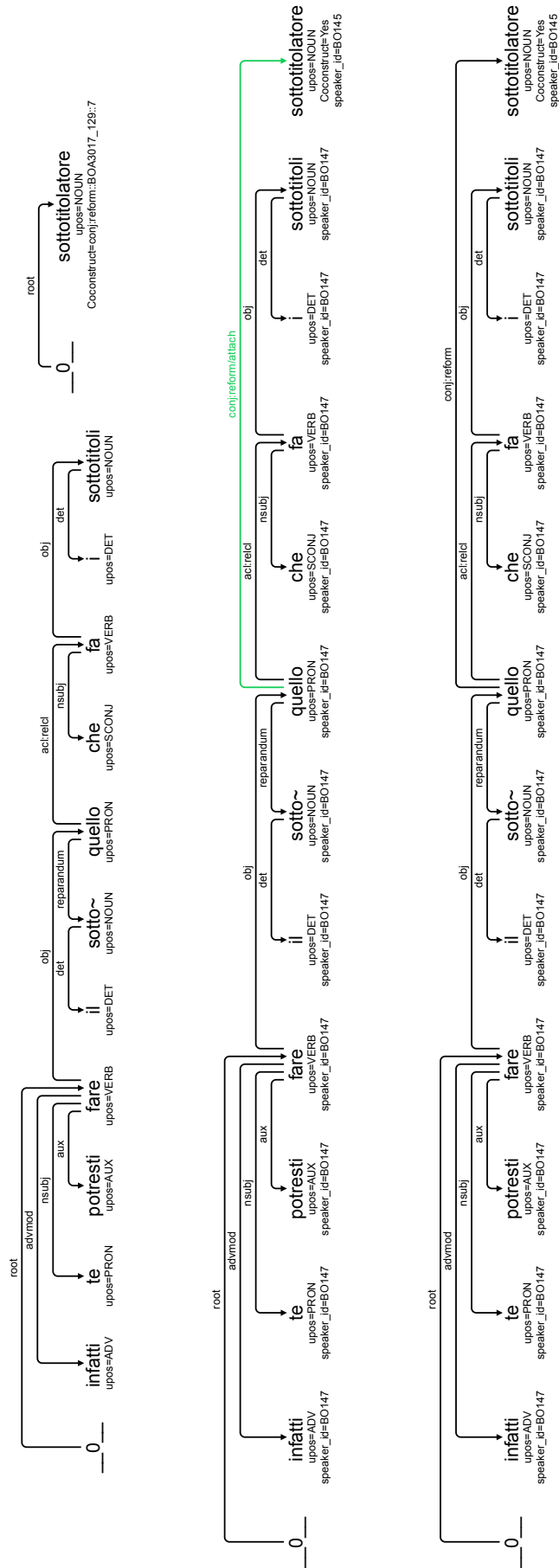


Figure 7: Italian repair example, see Example (12) for translation.