CGELBank Annotation Manual v1.0

Brett Reynolds, Nathan Schneider, and Aryaman Arora

Contents

Chapter 1

Introduction

CGELBank (Reynolds et al., 2023) is a treebank and associated tools based on a syntactic formalism for English derived from the *Cambridge Grammar of the English Language* (CGEL; Huddleston and Pullum, 2002).[1] It is hosted on GitHub at `https://github.com/nert-nlp/cgel`. This document lays out the particularities of the CGELBank annotation scheme.

As CGELBank is based on CGEL, CGEL itself should be the go-to resource for answers to most questions about the framework like "what is a determiner" or "what is the structure of the pseudo-cleft construction?" As much as possible, CGELBank follows the analysis set out in CGEL (and its companion textbook, *A Student's Introduction to English Grammar, 2nd edition* (Huddleston et al., 2021)). But that doesn't mean 100% correspondence. The purpose of this document, then, is to clarify the approach and highlight cases of non-correspondence. We see four general reasons why CGELBank and CGEL may not correspond precisely.

1. We believe CGEL has not been sufficiently explicit or exhaustive.
2. We believe CGEL has made an error.
3. We wish to represent something in a simpler or clearer way, even though the first two reasons do not apply.
4. We have made an error (please, let us know).

The most obvious case of being inexhaustive is in the lexicon. CGEL does not and could not list the category of every English word, nor would that be useful, as most dictionaries do a mostly fine job of identifying nouns and verbs. When it come to the closed classes, though, one might wonder whether *certain*, for instance, is always an adjective or whether it is sometimes a determinative. We attempt to provide exhaustive lists of pronouns, determinatives, prepositions, subordinators, and coordinators.

Another example of an ambiguity in CGEL, this one pertaining to syntax, is whether a relative pronoun in a relative clause like *who ate there* is a prenucleus followed by a subject gap or whether it is an in-situ subject. (We take the position that all relative clauses have gaps; see §4.5.1.) Where possible, we turn to subsequent publications by

---

[1] Page references in this document are to pages in CGEL unless otherwise noted.

CGEL authors (Payne et al., 2007, 2010, 2013; Pullum and Reynolds, 2013) to resolve ambiguities.

When it comes to errors, we believe there are only a very few in CGEL, and even those we are uncertain about. We believe, though, that the treatment of coordinates as unheaded, is unmotivated for reasons we explain in chapter 5.

An example (perhaps the only one) of a different representation – rather than a different analysis – relates to the way we portray supplements in tree structure, a topic also discussed in §3.1.3.

Chapter 2

The lexicon, categories, and functions

2.1 Introduction

Each lexeme is assigned a lexical category (e.g., noun, verb, adjective). These categories typically project phrases; for example an adjective projects and adjective phrase. The relations between phrases are labeled with functions (e.g., head, object, determiner; Ch. 3 discusses how this is realized in the trees[1]).

2.2 Categories

CGEL and CGELBank have the following lexical categories and phrasal projections:

Noun (N, N_{pro})	→ Nominal (Nom)	→ Noun phrase (NP)
Verb (V, V_{aux})	→ Verb phrase (VP)	→ Clause (incl. $Clause_{rel}$)
Adjective (Adj)		→ Adjective phrase (AdjP)
Adverb (Adv)		→ Adverb phrase (AdvP)
Preposition (P)		→ Preposition phrase (PP)
Determinative (D)		→ Determinative phrase (DP)
Interjection (Int)		→ Interjection phrase (IntP)[2]
Subordinator (Sdr)		
Coordinator		

Non-headed non-lexical categories (including Coordination and GAP) are presented in §2.2.3.

2.2.1 Lexical Projection Principle

Outside of morphologically derived expressions (§2.3.3), and excepting coordinators and subordinators, **a lexical node almost always projects a phrase of the corresponding category** shown in the table above. Thus, every N must serve as head within a Nom; every V must head a VP; every Adj must head an AdjP; and so forth.[3]

The one exception is that subject-auxiliary inversion (§4.3) targets auxiliaries specifically (rather than the VP they would project in normal position), so if the constituent in Prenucleus function consists of a single unmodified V_{aux}, it will not project a VP there.

2.2.2 Two-leveled phrases

As can be seen in the table above, nouns and verbs project two distinct levels of phrase structure. Nouns function as heads of Noms, which typically function as heads of NPs but which may also function as modifiers in a Nom, as in *a multiple choice question*.

Similarly, verbs function as heads of VPs, and VPs typically function as heads of Clauses, but they may also function as modifiers in a Nom, for instance *the quickly flowing water*.

---

[1] In this document, functions appear in a non-serif font like this: Head.

[2] See fn. 70 on p. 1361.

[3] The lexical head may be sister to a complement or modifier, if there is one (Ch. 3).

2.2.3 Non-headed categories

1. Coordination[4]
2. GAP (gap terminals are represented with an en-dash '–')
3. Nonce categories are written with the categories of the child constituents joined by the + operator (e.g., NP + PP for *the children in tow*).

2.2.4 Subcategories

Auxiliary verbs bear the label V_{aux}, pronouns bear the label N_{pron}, and relative clauses bear the label $Clause_{rel}$.[5]

Not clauses

The following sentences are main clauses in CGEL (p. 944) but not in CGELBank:

1. Clauses with subordinate form (e.g., *That it were true!*). These are subordinate clauses in CGELBank.
2. Conditional fragments (e.g., *If only I could!*). These are PPs in CGELBank.
3. Verbless directives (e.g., *Out of my way!* or *This way!*). These are XPs in CGEL-Bank.[6]
4. Parallel structures (e.g., *The sooner, the better!*). These are nonce XP + XP in CGELBank.)

    Subordinate verbless clauses such as *With <u>the kids in tow</u>, he headed out*, are treated as nonce constituents (here NP + PP).[7]

2.3 Lexemes

Where in doubt about the category of a given lexeme, consult the the Simple English Wiktionary. Note that determinatives are called "determiners" there.

2.3.1 Small categories

A mostly exhaustive list of each of the following categories is provided at the Simple English Wiktionary. Follow the links.
1. Prepositions
2. Coordinators
3. Subordinators
4. Determinatives (called "determiners" on the Wiktionary)

---

[4]The categories "coordinator" and "coordination", along with the function `coordinate` are always written out in full to limit confusion.

[5]Other clause types may be added in the future (§6.4).

[6]X is a variable for a lexical category, so an XP is a phrase ultimately headed by an X.

[7]*The kids in tow* may be a clause semantically, but a syntactic clause in CGEL is a projection of the VP. In a footnote on p. 1286, CGEL says, that "the ultimate head of *hat in hand* is *in*..., with *hand* an internal complement (*in hand* constituting the predicate) and *hat* an external complement (more specifically, the subject)." This is then a kind of 3rd layer on the PP analogous to the NP over the Nom or the Clause over the VP

2.3.2 Complex lexical items (written with a space)

CGELBank includes a small number of complex lexemes listed in (1), as follows: (a) determinatives, (b) prepositions, (c) subordinators, and (d) coordinators.

(1) a. *a certain*   *a few*   *a great many*   *a little*   *many a*   *no one*
      b. (i) *as for*   *as from*   *as if*   *as of*   *as per*   *as though*
               as to   *in case*   *in charge*   *in front*   *in order*   *in spite*
               in view   *no matter*   *on board*   *on purpose*   *on to*   *on top*
               so as   *à la*
         (ii)   *as long as* ("if")   *as soon as* ("when")
      c. *whether or not*   *whether or no*
      d. *as well as*   *rather than*

These items exhibit a high degree of grammaticalization, preventing their parts from being analyzed as syntactically separate units.

We take items such as *We plan to flight test and operate it* to be compounds (p. 1644–), not complex lexical items. They should be rewritten with a hyphen.

2.3.3 Morphologically derived complex expressions

Other expressions written as multiple words are treated as derived from a process of (productive) compounding that is more morphological than syntactic.

Flat nouns and determinatives

Certain expressions comprised of nouns but lacking ordinary headed NP structure are considered "flat" expressions (borrowing the terminology of the Universal Dependencies project; de Marneffe et al., 2021):

Multiword proper names (especially personal names: (2)), dates, and terminology (*carbon dioxide*) often fall into this category, along with enumerated nouns like *page 27*

or *building J* (p. 518).[8] Numeric expressions comprised of determinatives (*120 million*: (3)), similarly form flat determinatives.[9] In these cases—and these cases only—there is a lexical category that is not a preterminal, but has lexical categories as children, each bearing the function Flat.

This is in contrast to proper names derived from ordinary NPs (e.g., *Yanhee Hospital*, *No Time to Die*), which receive ordinary NP analyses.

Zero-derived NPs

Proper names may take the form of a phrase other than an NP—for example, a title of a book that takes the form of a Clause, VP, or PP (but is treated as an NP with respect to the rest of the sentence). In such cases, we show the internal syntactic structure of the source phrase at the bottom of the tree, and then the phrase as a whole is reanalyzed as an NP via the Compounding function,[10] as in (4). If the expression is made of disjoint recognizable syntactic constituents, each attaches as Compounding, as in (5).[11]

(4)

---

[8]However, formulations like *buildings J and K* suggest some of these idiosyncratic name patterns may call for headed analyses, though the head is not always obvious (Schneider and Zeldes, 2021). This requires further investigation.

[9]This includes numbers such as *twenty seven*, *two thousand three*, *three oh one*, etc., as well as times like *three fifteen*.

[10]Not to be confused with *syntactic* nominal compounds, which feature an NP with an internal Mod.

[11]In the future, this strategy could be extended to syntactically anomalous idioms that are not NPs, like *Long time no see*. We have not encountered such cases in the corpus.

(5)

```
                              VP
              ┌───────────────┴───────────────┐
           Head:                            Obj:
            V                                NP
            │                                │
            │                              Head:
            │                              Nom
            │              ┌────────────────┼────────────────┐
            │        Compounding:      Compounding:      Compounding:
            │            P                 PP                NP
            │            │             ┌────┴────┐      ┌────┴────┐
         reading        If       on a Winter's Night    a Traveler
```

2.4 Functions

CGELBank uses the subset of the functions used in CGEL showing in Figure 2.1. Notably, CGELBank uses Head for CGEL's Predicate and Predicator.

2.4.1 Modifier vs. Supplement

CGEL's description of adjuncts is not entirely clear about the division between modifiers and supplements in clause structure (see, for instance, similarities highlighted on p. 1360). As a default, we take adjuncts of time, place, manner, condition, reason, and so on to be modifiers—even when offset by a comma. The function of supplement should be reserved for those adjuncts which are quite clearly presented as addenda: speaker commentary, clarification, parentheticals, background descriptions, and the like.

(6) Modifier

    a. ***Because it was raining***, *I was reluctant to go outside.*
    b. ***If you're hungry***, *have a sandwich.*
    c. ***When the president stands***, *nobody sits.*

See §3.1.2 regarding the branching structure of modifiers. Note that post-head modifiers in a clause generally attach to the lowest VP.

(7) Supplement

    a. ***As has become clear***, *Sharon cannot be trusted.*
    b. *You are recommended to get a flu shot, **especially if you are over 50**.*
    c. ***Given the size of the task***, *I think we can expect it to take a long time.*
    d. ***Well***, *I think I'll just stay here.*
    e. ***Legally***, *he's too young.*
    f. ***Remarkably***, *we were not late.*
    g. ***Frankly***, *I wouldn't.*

Figure 2.1: Taxonomy of functions. The labels that appear in CGELBank are in **bold-sans-serif**.

h. *There is, **however**, a catch.*
   i. ***Damn**, what was I thinking?*
   j. *What are you up to, **Kate**?* (Vocative, a subtype of supplement)
   k. *I did it, **which was a good move**.* (relative clause as supplement)
   l. *I did it, **clearly a good move**.*

2.4.2   Modifier vs. Complement

CGEL (starting on p. 439) regards some pre-head dependents in nominals as comple-
ments rather than attributive modifiers, as in *a <u>flower</u> seller* or *an <u>income tax</u> adviser*. In
contrast, we analyze these pre-head dependents as modifiers (consistent with Huddleston
et al., 2021, p. 128).

2.4.3   Nonce functions

Where a coordination has nonce constituents, we assign a composite function consisting
of the function of each constituent within the coordinate concatenated with +. For
example, in *I gave <u>$10 to Kim</u> and <u>$5 to Pat</u>*, *$10* and *$5* are objects, and *to Kim* and *to
Pat* are complements, so the function of the coordination is Obj+Comp. In some cases,
where the functions are not be consistent across coordinates, we give them with a slash
notation, as in Obj+PredComp/Comp, indicating that the first coordinate includes Obj and
PredComp and the second Obj and Comp.

Chapter 3

Tree structure and style

3.1 Branching and labeling basics

We use the term "tree" here loosely to mean a syntax tree. Strictly, the trees are directed acyclic graphs (DAGs; see §3.2), and, for the most part they are true trees as defined in graph theory. The basic rules of branching in CGEL are 1) that phrases, along with non-headed constituents (see §2.2.3), can be represented by phrase-structure trees, 2) that trees are generally right branching, 3) that binary branching is preferred, and 4) that *n*-ary branching is possible. Phrases-structure trees are constructed by breaking a phrase into its constituents, which are, in turn, represented as phrase-structure trees.

Most nodes are labeled with both a category and a function. These functions being relational, they apply to the incoming branch (edge), as in (8) but are displayed at the node for convenience, as in (9). The top node on the tree has a category label only. The terminal nodes are words or, in a few cases, an en-dash '–' representing a gap.

(8)
VP
Head — V | *eat*
Object — NP △ *apples*

(9)
VP
Head: V | *eat*
Obj: NP △ *apples*

3.1.1 Unary branching

A clause like *stop* is headed by a VP, which is headed by a V. This is an example of unary branching. CGEL often omits phrasal nodes in unary branches. For example, the determiner in *some children* is represented as a D instead of a DP in CGEL's [13] (p. 26), and no Nom is shown between the NP and the N. This tree from CGEL is reproduced in (10).

(10)
NP
Det: D | *some*
Head: N | *children*

In CGELBank, nodes are never omitted, so that we represent the same NP with unary branches, as in (11).[1]

---

[1] In this guide, we sometimes omit internal structure, but this is always indicated by a triangle in the tree.

(11)

```
                NP
         ┌───────┴───────┐
       Det:            Head:
        DP             Nom
        │               │
      Head:           Head:
        D               N
        │               │
      some           children
```

A single modifier may be the sister of a lexical head. This is exemplified in (12), where each modifier (which may itself be a Nom constituent) attaches at a Nom. The modifier nearest to the head shares the Nom projected by the head, while remaining modifiers add extra Nom layers. Note that per the Lexical Projection Principle (§2.2.1), each modifier lexeme projects its own Nom level, which is never shared with the head.

(12)

```
                     NP
                     │
                   Head:
                    Nom
            ┌─────────┴─────────┐
          Mod:               Head:
          Nom                 Nom
           │             ┌──────┴──────┐
         Head:          Mod:        Head:
           N            Nom           N
           │             │            │
                       Head:          │
                         N            │
                         │            │
        desert        weather      stations
```

In (11, 12), for every word except *stations*, the phrasal constituent projected by the lexical node is unary. Likewise, an intransitive unmodified V may be the sole element of a VP. Coordinations of intransitive unmodified verbs must bear the unary VP layer as well (even in the head of a marked coordinate; note that coordination in CGELBank is always between phrasal categories: §5.2.2).

Every clause must be headed by a VP (or another clause level), and every NP must be headed by a Nom (or another NP level). Clause and NP constituents may be unary to conform to this requirement.[2]

---

[2]The term "clause" here covers constituents of both the Clause and Clause$_{rel}$ categories.

16

3.1.2 Binary branching

Most headed nonlexical constituents exhibit binary branching. If a phrase of type XP contains multiple dependents, these will generally be layered so as to attach one at a time going outward from the head, forming intermediate XP constituents. A typical example is (13), where the VP-internal complement (Obj) forms a VP constituent with the head verb, and that VP heads a larger VP constituent with a modifier.

(13)

```
                    VP
             ┌──────────────┐
           Mod:           Head:
           AdvP            VP
            △         ┌─────────┐
                    Head:     Obj:
                      V        NP
                      │        △
         quickly     eat     apples
```

Exceptions to binary branching are discussed in §3.1.3.

Modification

Modification is always binary: a constituent attaching as Mod always has Head as its sister.[3] A modifier will never be sister to a complement, for example.

Where there are multiple modifiers, these are layered, one per XP. Generally those farther from the head are higher in the tree structure, but semantics are also brought to bear in adjudicating when there are both pre- and post-head modifiers.

Post-head modifiers attach as low as possible in the tree structure. For example, in principle, *quickly* in *It can run quickly* could attach in the *can* clause, the *can* VP, the *run* clause, or the *run* VP. Unless there is a clear reason to do otherwise, we attach it to the *run* VP.

One upshot of this preference for binary branching is that, while it is true in principle that, for example, a VP headed by a transitive verb licenses an object and permits certain kinds of modifiers, it is not the case that a modifier or complement may always be branched from any VP. This also applies to an XP coordinate with a marker (see §3.1.3).

Post-head pre-complement modifiers

Should a modifier come between a head and its complement, there are two possible analyses. The first is that the complement is post-posed (see §4.2.2), usually because the modifier is heavy. In such a case, we include a gap in the immediate post-head position in an XP with the head. The modifier is attached to a higher XP, and the post-posed complement to a yet higher XP.

---

[3]Binary not counting supplements: §3.1.3.

In some cases, though, a modifier naturally comes between a head an its complement, so the inner VP will contain the modifier whereas the outer VP will add the complement. We see this most commonly with auxiliary verbs (e.g., *You may not go*). In such cases, we take *not* to form a VP with *may*, and this VP takes the clause *go* as its complement. A non-auxiliary example attested in CGELBank is *think again about. . .*, where the *about*-PP is licensed by the verb.

3.1.3   Beyond binary branching

Morphologically derived expressions

Complex expressions such as personal names may contain two or more Flat or Compounding dependents: see §2.3.3.

Coordinations

Coordination constituents are non-headed and may have more than two daughters in Coordinate function. See Ch. 5.

VP-internal complements

While multiple modifiers are layered, multiple (internal, non-extraposed) complements within a VP will typically attach on the same level. A typical example of ternary branching would be in a ditransitive construction like (14).

(14)

$$\text{VP}$$

Head:	Obj$_{ind}$:	Obj$_{dir}$:
V	NP	NP
give	*us*	*the apples*

Extraposed complements trigger a separate level: see §3.4.4.

The treatment of complements within VPs is exceptional. In NP structure, complements and modifiers are treated the same: each attaches within a separate Nom layer.

Supplements

CGEL analyzes supplements as not being fully integrated into phrase structure. To illustrate this, it shows them as separate trees with an arrow pointing to the supplement's semantic anchor as in (15) (CGEL's [12], p. 1354).

Clause ◄- Supplement:
NP

Subject   Predicate:
NP   VP

Jill   *sold her internet shares in January*   *a very astute move*

Without disagreeing with the analysis in CGEL, CGELBank presents supplements as an additional branch from the anchor for simplicity, as in (16).

(16)   Clause

Subj: Head: Supplement: Supplement:
NP   VP   NP   AdvP

Jill   *sold her internet shares in January*   *a very astute move*   *frankly*

3.1.4 Summary of constraints on branching

Putting together the branching principles expressed above and in the Lexical Projection Principle (§2.2.1), and setting aside non-headed constituents, we have the following constraints:

1. A node of a lexical category that projects a constituent of a corresponding phrasal category must do so (as its head) unless it is V_{aux} in Prenucleus function.

2. A lexical node has too many layers if its parent is in Head function, the categories of its parent and grandparent are the same, and the grandparent is not functioning as a Coordinate.

3. In a VP, constituents in non-extraposed internal complement[4] functions will be on the same level except where separated from the head by an intervening modifier. Every modifier and extraposed complement forms a binary VP with the head (not counting supplements).

    In other words, each non-unary VP level consists of a Head plus a) a single Mod, b) an extraposed complement, or c) any number of non-extraposed internal complements (and there should not be two consecutive levels of type (c)).

4. An NP must be headed by a Nom or another NP level.

5. A Clause or Clause$_{rel}$ must be headed by a VP or another Clause or Clause$_{rel}$ level.

6. A phrasal (headed) constituent other than VP must be no more than binary (not counting supplements).

7. A unary phrase (not counting supplements) should not be headed by a constituent with the same phrasal category as this would be a vacuous branch—though an exception is necessary for the intermediate node of fused structures described below.

---

[4]Non-extraposed internal complements are constituents in the function of Comp, Obj, Obj$_{dir}$, Obj$_{ind}$, Pred-Comp, Particle, or DisplacedSubj. CGEL makes a further distinction of *core* vs. *non-core* internal complements, but this distinction is not currently reflected in CGELBank.

3.2 Fusion of functions

CGEL departs from strict tree structure in cases where a constituent functions both as a head and as a dependent while still being a directed acyclic graph (DAG), as detailed in (Pullum and Rogers, 2009). A constituent has two parent constituents with respect to which it bears different functions (one of them head); these functions are shown hyphenated in a single label. Typically, the constituent with two parents is deeper in the tree by one level along one branch than along the other, though the difference may involve multiple levels as in (19) below.

3.2.1 Determiner-Head

(17)

NP

Head:
Nom

Det-Head:
DP

this

(18)

NP

Head:
Nom

Det-Head:
DP

Mod:
AdjP

hardly anyone   *present*

(19)

NP

Head:
Nom

Head:
Nom

Mod:
Clause$_{rel}$

Det-Head:
DP

Mod:
AdjP

everyone   *present*   *that I knew*

(20)

NP

Head:
Nom

Det-Head:
DP

Comp:
PP

Head:
P

Obj:
NP

some   *of*   *those*

3.2.2 Modifier-Head

(21)

```
                    Clause
             ┌────────┴────────┐
          Subj:             Head:
           NP                VP
       ┌────┴────┐
     Det:      Head:
      DP        Nom
                  ╲
               Head:
                Nom
                 │
              Mod-Head:
               AdjP

      the      rich     get richer
```

3.2.3 Head-Prenucleus

The Head-Prenucleus function appears in fused relative constructions. (For non-fused relatives, see §4.5.1.)

Relative NP

(22)

Relative PP

Though relative clauses do not typically function as modifiers in PPs, this is the analysis in CGEL and also in Payne et al. (2007).

(23)

PP

Mod:
Clause$_{rel}$

Head-Prenucleus:
PP$_x$

Head:
Clause$_{rel}$

Subj:
NP

Head:
VP

Head:
V$_{aux}$

Comp:
GAP$_x$

wherever   *you*   *are*   –

Relative AdjP

(24)

AdjP

Mod:
Clause$_{rel}$

Head-Prenucleus:
AdjP$_x$

Head:
Clause$_{rel}$

Subj:
NP

Head:
VP

Head:
V$_{aux}$

PredComp:
GAP$_x$

however small   *it*   *is*   –

23

Relative AdvP

(25)

AdvP
- Head-Prenucleus: AdvP$_x$
  - *however often*
- Mod: Clause$_{rel}$
  - Head: Clause$_{rel}$
    - Subj: NP
      - *I*
    - Head: VP
      - Head: V — *try*
      - Mod: GAP$_x$ — *–*

3.3 Indirect complements

An indirect complement forms a binary branch with the lowest possible XP. In (26), the Comp$_{ind}$ is licensed by *so*.

(26)

NP
- Head: NP
  - Mod: AdjP — *so great*
  - Head: NP — *a loss*
- Comp$_{ind}$: Clause — *that we gave up*

See §4.5.3 for examples of indirect complements with gaps.

3.4 Information packaging clause constructions

3.4.1 Pre- and Postposing

CGEL claims that "subject postposing affects order, not function" (p. 244), but this is not strictly true, at least not in the CGELBank tree structure. The pre- or postposed constituent is co-indexed to a gap with the function the constituent would have had in the basic position, but the constituent itself is a pre- or postnucleus.

(27)

```
                                      Clause
                 ┌──────────────────────┴──────────────────────┐
            Head:                                         Postnucleus:
            Clause                                             NP
        ┌─────┴─────┐                                          ╱╲
     Subj:        Head:                                       ╱  ╲
      NP            VP                                       ╱    ╲
      ╱╲      ┌─────┼─────┐                                 ╱      ╲
     ╱  ╲   Head:  Obj:  Comp:                             ╱        ╲
    ╱    ╲    V    GAP_X   PP                              ╱          ╲
    │     │   │     │      ╱╲                             ╱            ╲
    he   gave  –   to charity    everything he had earned from the years of toil
```

Additional examples appear in §4.2.

3.4.2 Inversion

The auxiliary verb is a Prenucleus co-indexed to a gap in the usual location. If the inversion is triggered by a fronted element, it is also a Prenucleus.

(28)

```
                              Clause
                ┌───────────────┴───────────────┐
          Prenucleus:                        Head:
          AdvP[x]                            Clause
                             ┌─────────────────┴─────────────┐
                       Prenucleus:                         Head:
                       V_aux[y]                            Clause
                                          ┌─────────────────┴──────────┐
                                        Subj:                        Head:
                                         NP                           VP
                                                        ┌──────────────┴──────────┐
                                                      Head:                      Comp:
                                                      GAP[y]                     Clause
                                                                                   │
                                                                                 Head:
                                                                                  VP
                                                                        ┌──────────┴──────────┐
                                                                      Head:                  Mod:
                                                                        V                   GAP[x]
                                                                        │                     │
          thus        had        they        –        parted          –
```
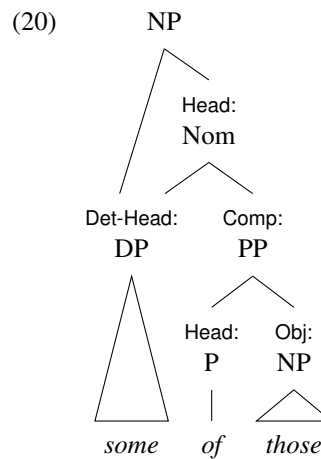
Additional examples and discussion appear in §4.2.3.

3.4.3 Existential clauses

The displaced subject is a complement in the VP along with any others such as a PredComp.

(29)

3.4.4 Extraposition

In CGELBank, extraposed subject is located in an outer VP layer (in line with Huddleston et al., 2021, p. 372):

(30)

3.4.5 Cleft clauses

The relative clause appears in extranuclear position at the end (p. 1416). Note that this postnucleus is not coindexed.

(31)

Clause
├─ Head: Clause
│ ├─ Subj: NP — *it*
│ └─ Head: VP
│   ├─ Head: V_{aux} — *was*
│   └─ PredComp: NP_x — *a bee*
└─ Postnucleus: $Clause_{rel}$
   ├─ Marker: Sdr — *that*
   └─ Head: Clause
      ├─ Subj: GAP_x — *–*
      └─ Head: VP — *stung me*

Occasionally, a nonfinite clause will appear instead of a relative clause (p. 1420). Without a relative clause, there is no gap in the postnucleus. Here is one in an interrogative sentence:

(32)

Clause
├─ Prenucleus: $V_{aux\,\boxed{x}}$ — *was*
└─ Head: Clause
   ├─ Head: Clause
   │ ├─ Subj: NP — *it*
   │ └─ Head: VP
   │   ├─ Head: $GAP_{\boxed{x}}$ — *–*
   │   └─ PredComp: NP — *a bee*
   └─ Postnucleus: Clause
      └─ Head: VP — *making all that noise*

3.4.6 Passives

Any internalised complement is a Comp in the VP.

(33)

Clause
├─ Subj: NP — *it*
└─ Head: VP
   ├─ Head: V$_{aux}$ — *was*
   └─ Comp: Clause
      └─ Head: VP
         ├─ Head: V — *followed*
         └─ Comp: PP — *by a comma*

3.4.7 Dislocation

The dislocated constituent is a Supplement.

(34)

Clause
├─ Supplement: NP — *me*
├─ Subj: NP — *I*
└─ Head: VP — *wouldn't do it*

Chapter 4

Gaps and co-indexing

4.1 Gapping basics

Formal syntactic theories have been known to postulate elaborate inventories of null elements and systems of movement in order to account for phenomena such as ellipsis and control. CGEL positions itself as a descriptive framework, and as a rule, avoids invisibilia—but unbounded dependencies and other noncanonical word order constructions are the exception. For such constructions, a **gap** node appears as a leaf in the tree to indicate the canonical position of a constituent, and the gap is coindexed with the overt constituent in its "surface" position. This makes for a fairly intuitive description of sentences with relative clauses, WH-questions, inversion, and pre-/postposed elements.

CGEL notates gaps with a '__', but we use an en-dash '–'. Gap nodes are labeled in the typical manner with a function and with "GAP" appearing as the category. Gaps are co-indexed with a subscript such as x to any extranuclear node or, where no such node exists, to an antecedent. A typical example is shown in (35).

(35)

A gap may even appear in extra-nuclear material, as in GAP_y in (36) from Mark Steedman.

(36)

Clause
- Prenucleus: Clause$_x$
  - Prenucleus: NP$_y$ — *whose woods*
  - Head: Clause
    - Subj: NP — *these*
    - Head: VP
      - Head: V$_{aux}$ — *are*
      - PredComp: GAP$_y$ — –
- Head: Clause
  - Subj: NP — *I*
  - Head: VP
    - Head: V — *think*
    - Comp: Clause
      - Subj: NP — *I*
      - Head: VP
        - Head: V — *know*
        - Comp: GAP$_x$ — –

A gap can also be co-indexed to a gap. *Which$_x$ was it $-_x$ [she put it in $-_x$]?*

Formal constraints. In CGELBank, we enforce the following requirements:

- Every gap must be coindexed to exactly one overt element in the sentence (and potentially to other gaps).
  - The overt element should bear a nonlexical category where possible. For example, interrogative or relative *who* should be coindexed at the NP level. Exceptions where coindexation is at the lexical level: the V$_{aux}$ in Prenucleus function with subject-auxiliary inversion (40); and an N that is sister to a bare relative clause modifier (44).
- Every distinct coindexation variable in the sentence must apply to at least one gap.
- A gap must have at least one non-gap sister in a function other than supplement (i.e. it must never be the child of a unary rule, and a binary rule must not consist simply of two gaps).
- No overt element or gap may receive more than one coindexation variable (e.g., a noun with two non-WH relative clause modifiers will have the two coindexation variables at different Nom levels).

Thus, constituents involved in coreference phenomena do not receive coindexation except to resolve a gap. This rules out (in most cases) ordinary anaphora, reflexive anaphors, resumptive pronouns, and nouns that are relativized with an overt relative pronoun (§4.5.1).

4.2 Pre- and postposing

A pre- or postposed dependent is indicated with a gap in the basic position co-indexed to the pre- or postposed element.

4.2.1 Preposing

(37)

4.2.2 Postposing

(38) Postposing due to weight (pp. 1382–1383):

(39) Subject postposing with verb of reporting (p. 1384):

```
                          Clause
              ┌─────────────┴──────────────┐
        Prenucleus:                      Head:
         Clauseᵧ                        Clause
                                 ┌─────────┴─────────┐
                              Head:             Postnucleus:
                             Clause                 NPₓ
                         ┌─────┴─────┐
                       Subj:       Head:
                       GAPₓ         VP
                                 ┌───┴───┐
                              Head:    Comp:
                               V       GAPᵧ
                               │         │
       it works    –    said    –    Jim
```

4.2.3 Adjuncts in clause structure

In CGELBank, pre-clause adjuncts are not said to be preposed – and thus do not trigger a gap[1] – except where the adjunct triggers subject-auxiliary inversion as in the following cases:

- A negative polarity item (e.g., _Never had I seen the like._)
- Phrases starting with _only_ (e.g., _Only once had I seen it._)
- Other adjuncts, such as _thus_ or _yet_ (e.g., _Thus had they parted._)

See §3.4.2 for an example tree.

4.3 Subject–auxiliary inversion

Subject–auxiliary inversion (SAI) is a gapped construction with a co-indexed element (typically a sole V_{aux}) in the prenucleus. This applies equally to _do_ support. A typical example is shown in (40).

---

[1]Contrast CGEL's _If you pay me_ᵢ, _I'll do it_ ₋ᵢ (p. 1092).

(40)

Clause

Prenucleus:
$V_{aux\ \boxed{x}}$

Head:
Clause

Subj:
NP

Head:
VP

Head:
$GAP_{\boxed{x}}$

PredComp:
Adj

were *you* *–* *okay*

Note that the prenucleus in SAI has no VP, just a single lexical item. In rare circumstances, a larger constituent must be created for the prenucleus:

(41)

Clause

Prenucleus:
Coordination$_x$

Head:
Clause

Subj:
NP

Head:
VP

Head:
GAP_x

Comp:
Clause

will or won't *you* *–* *do it*

(See (53) for the structure of the coordination of auxiliaries.)

4.4   Subject–dependent inversion

Subject–dependent inversion (SDI) is a double-gapped construction with a subject gap co-indexed to the postnucleus and another gap (typically a complement of the head verb) co-indexed to the prenucleus (p. 1385). A typical example is shown in (42).

(42)

```
                              Clause
                   ┌────────────┴────────────┐
              Prenucleus:                  Head:
                 PP_y                     Clause
                              ┌─────────────┴─────────────┐
                           Head:                      Postnucleus:
                          Clause                          NP_x
                   ┌────────┴────────┐
                 Subj:            Head:
                GAP_x              VP
                           ┌────────┴────────┐
                         Head:             Comp:
                           V              GAP_y

      here      –      is        –       Jim
```

4.5   Unbounded dependencies

4.5.1   Relative constructions

Every relative construction includes a clause with a gap that is co-indexed to a prenucleus or antecedent. This includes subject relatives such as (43).

(43)

```
                     NP
           ┌──────────┴──────────┐
        Det:                   Head:
         DP                     Nom
                       ┌──────────┴──────────┐
                    Head:                   Mod:
                    N_pro                 Clause_rel
                                   ┌──────────┴──────────┐
                             Prenucleus:              Head:
                                NP_x               Clause_rel
                                             ┌────────┴────────┐
                                           Subj:            Head:
                                          GAP_x              VP

      a     person     who        –       works
```

Note that, as stipulated in §4.1, the gap is coindexed to just one overt antecedent—in this case, the relative pronoun *who*. The pronoun's antecedent (*person*) is therefore not coindexed.

In *that* relatives and bare relatives, however, there is no extranuclear constituent, and so the co-indexation is with the antecedent, as in (44, 45).

(44) Bare relative:

```
                        NP
               _____
              |                   |
           Det:                Head:
            DP                  Nom
            |            _____
            |           |                |
            |        Head:             Mod:
            |         Nₓ            Clause_rel
            |          |         _____
            |          |        |              |
            |          |      Subj:          Head:
            |          |       NP             VP
            |          |        |         _____
            |          |        |        |         |
            |          |        |     Head:      Obj:
            |          |        |      V        GAPₓ
            |          |        |      /\         |
            a       person    you   know          –
```

(45) *That* relative:

37

```
                          NP
              ┌────────────┴────────────┐
           Det:                      Head:
            DP                        Nom
                              ┌────────┴────────┐
                           Head:             Mod:
                            N_x           Clause_rel
                                        ┌─────┴─────┐
                                     Marker:      Head:
                                       Sdr      Clause_rel
                                              ┌────┴────┐
                                           Subj:      Head:
                                            NP         VP
                                                   ┌────┴────┐
                                                Head:      Obj:
                                                  V       GAP_x
            a      person      that      you     know       –
```

See also fused relatives, §3.2.3.

4.5.2   Open interrogative clauses

Every open interrogative clause has a gap co-indexed to an interrogative phrase as in
(46) unless the there is no inversion, in which case the interrogative phrase is in situ as
in *Who did that?* or *You did what to him?*. This contrasts with relative clauses in which
there is always a gap, even in subject function.

(46)

Clause

Prenucleus:
NP_x

Head:
Clause

Prenucleus:
$V_{aux\,y}$

Head:
Clause

Subj:
NP

Head:
VP

Head:
GAP_y

Comp:
Clause

Head:
VP

Head:
V

Obj:
GAP_x

who   *did*   *you*   *–*   *see*   *–*

4.5.3 Hollow clauses

Every hollow clause has a gap co-indexed to an antecedent.

(47)

Clause

Subj:
NP_x

Head:
VP

Head:
V_{aux}

PredComp:
AdjP

Head:
AdjP

$Comp_{ind}$:
Clause

Mod:
AdvP

Head:
AdjP

Head:
VP

Marker:
Sdr

Head:
VP

Head:
V

Obj:
GAP_x

the box *was* *too* *heavy* *to* *lift* *–*

40

(48)

```
                              NP
                  ┌───────────┴───────────┐
               Det:                     Head:
               DP                        Nom
                │              ┌──────────┴──────────┐
                │           Head:                 Comp_ind:
                │           Nom_x                   Clause
                │        ┌────┴────┐                  │
                │      Mod:     Head:              Head:
                │      AdjP       N                 VP
                │       │         │          ┌───────┴───────┐
                │       │         │       Marker:         Head:
                │       │         │        Sdr             VP
                │       │         │          │        ┌─────┴─────┐
                │       │         │          │      Head:       Obj:
                │       │         │          │        V        GAP_x
                a     tough      box        to       lift        –
```

Note that in (48), the gap is coindexed with the ʜᴇᴀᴅ Nom *tough box*, even though strictly speaking the adjective phrase is not understood as part of the gapped material.

4.6 Coordination

A number of non-basic coordination constructions include gaps. For examples, see §5.3.

4.7 Non-gapped constructions

The following structures have at least a notional gap, but any such gaps are not annotated in CGELBank.

4.7.1 Comparatives

CGEL writes some examples of comparative clauses with a gap, as in *It is as deep as* [*it is ___ wide*], but we view this as heuristic and do not include any such gaps in tree structure. CGELBank does, however, mark such clauses with searchable but non-visible features.

4.7.2 Subjects in infinitival and participial clauses

CGELBank does not mark a gap for a subject in infinitival clause or participial clause, whether or not there is control or raising. As a result, the gaps and co-indexation below will not appear in the tree structure, regardless of whether they are controlled as in (49) or not, as in (50).

(49) a. *I_x hope __$_x$ to see her.*
     b. *They asked Pat_x __$_x$ to help them.*

(50) a. *That would mean __ starting over.*
     b. *She_x has an invitation __$_x$ to attend.*

4.7.3 Ellipsis

Similar to the case in comparative clauses, we view CGEL's "gaps" indicating ellipsis to be heuristic. For treatment of trees with ellipsis, see §6.1.

4.8 Co-indexing without gaps

CGELBank does not co-index other anaphora, including displaced or extraposed constituents (but see §4.2).

Chapter 5

Coordination

5.1 Introduction

CGELBank generally follows CGEL in its analyses, but, consistent with SIEG2, we are explicit in taking coordinates to be headed phrases and markers to be dependents therein. This parallels CGEL's treatment of markers in subordination (9, p. 954).

Also, CGEL labels coordinations with the category of coordinates (e.g., NP-coordination). In contrast, CGELBank simply labels all coordinations as "Coordination".

Here, we show how this difference in analysis affects the trees in CGELBank, while, at the same time, setting out our analysis of a number of points about which CGEL is inexplicit.

We also treat supplementation here, not because we disagree with CGEL's analysis, but because supplements, like coordinate phrases, can have a coordinator as a marker and because, for simplicity, we deviate from CGEL's style of indicating supplements with an arrow pointing to the anchor in the tree (see 12 on p. 1354).

5.2 Basic coordination

The most basic cases of coordination have two phrasal coordinates with a coordinator functioning as a marker in the second coordinate as in (51).

(51)

5.2.1 Asyndetic coordination

Asyndetic coordination, as in (52), lacks a marker such as *and*. Most of what is presented here will deal explicitly with syndetic coordination, but will almost always apply equally to asyndetic coordination, except, of course, for discussion of markers.

(52)

Clause
Head:
VP
Head:
V
Obj:
Coordination
Coordinate:
NP
Coordinate:
NP
Coordinate:
NP
bring *games* *stories* *songs*

5.2.2 Coordination of lexemes

CGELBank treats apparent coordination of lexemes as coordination of phrases, as in (53). (Appendix A presents an in-depth comparison of alternatives to justify this approach.) The one exception to this is discussed in §5.4.2.

(53)

VP
Head:
Coordination
Comp:
Clause
Coordinate:
VP
Coordinate:
VP
Head:
V$_{aux}$
Marker:
Coordinator
Head:
VP
Head:
V$_{aux}$
can *and* *will* *try*

(54)

NP

Head:
Nom

Head:
Coordination

Comp:
PP

Coordinate:
Nom

Coordinate:
Nom

Head:
N

Marker:
Coordinator

Head:
Nom

Head:
N

development   *and*   *implementation*   *of policy*

(55)

AdjP

Mod:
AdvP

Head:
Coordination

Coordinate:
AdjP

Coordinate:
AdjP

Head:
Adj

Marker:
Coordinator

Head:
AdjP

Head:
Adj

very   *friendly*   *and*   *helpful*

5.2.3 Root-branched coordinators and coordinates

Here we consider cases in which there is a sentence-initial coordinator or where the root is a coordination. (The related case of a supplement marked with a coordinator will be dealt with in §5.4.5.)

Where a sentence starts with a coordinator, we give that sentence its usual constituent label, so that (56) *And so it begins* is a clause, and (57) *But not that* is an NP.

(56)
```
                    Clause
                  /        \
            Marker:        Head:
          Coordinator      Clause
               |         /        \
             and       so it begins
```

(57)
```
                     NP
                  /      \
            Marker:      Head:
          Coordinator     NP
               |        /      \
             but       not that
```

When the sentence consists of two or more coordinates, the root is a coordination, not a clause or other XP, as in (58).

(58)
```
                 Coordination
               /              \
        Coordinate:        Coordinate:
          Clause             Clause
            |              /         \
            |          Marker:       Head:
            |        Coordinator      Clause
          stay           |          /       \
                        and       have some coffee
```

5.2.4 Layered coordination

(59)

Coordination

- Coordinate: Coordination — *small and quiet*
- Coordinate: Coordination
  - Marker: Coordinator — *but*
  - Head: Coordination
    - Coordinate: AdjP — *artful*
    - Coordinate: AdjP
      - Marker: Coordinator — *and*
      - Head: AdjP — *enterprising*

5.2.5 Correlative coordination and marker category

A DP may function as marker in the first coordinate of a coordination in correlative coordination. (See also CGEL's [45] on p. 1308 for an example of a gapped marker.)

(60)

Coordination

- Coordinate: AdjP
  - Marker: DP — *neither*
  - Head: AdjP — *artful*
- Coordinate: AdjP
  - Marker: Coordinator — *nor*
  - Head: AdjP — *enterprising*

A subordinator may also appear in marker function, so that it's possible to have strings of two markers like *either to* and *or to* in (61).

(61)

```
                              Coordination
                    _____|_____
              Coordinate:                      Coordinate:
                  VP                               VP
           _____|_____                 _____|_____
       Marker:        Head:            Marker:          Head:
         DP             VP           Coordinator          VP
          |         ____|____             |          ____|____
         /\     Marker:   Head:           |      Marker:   Head:
        /  \     Sdr       VP             |       Sdr       VP
       /    \     |        /\             |        |        /\
     either  to  live                    or       to    let live
```

5.2.6   Expansion of coordinates by modifiers

The structure in CGELBank differs from CGEL's (see CGEL's [9] on p. 1278). Here, the AdvPs in an NP like *the guests and indeed his family too* from (62) are analyzed as peripheral modifiers or possibly, in the case of *indeed*, as supplements.

(62)

```
                         Coordination
                 _____|_____
           Coordinate:                  Coordinate:
              NP                            NP
              |            _____|_____
             /\        Marker:         Supplement:        Head:
            /  \     Coordinator          AdvP             NP
           /    \         |                |           _____|_____
          /      \        |               /\       Head:       Mod:
         /        \       |              /  \        NP         AdvP
     the guests   and            indeed      his family    too
```

5.3   Non-basic coordination

5.3.1   Right nonce-constituent coordination

CGEL gives no function to a right nonce-constituent coordination (p. 1342), but to avoid this functionless node, CGELBank uses the + operator to create a nonce function, as in (63).

(63)

Clause
├─ Subj: NP — *Mo*
└─ Head: VP
    ├─ Head: V — *gave*
    └─ Obj_ind + Obj_dir + Mod: Coordination
        ├─ Coordinate: NP + NP + PP
        │   ├─ Obj_ind: NP — *me*
        │   ├─ Obj_dir: NP — *one*
        │   └─ Mod: PP — *before*
        └─ Coordinate: NP + NP + PP
            ├─ Marker: Coordinator — *and*
            └─ Head: NP + NP + PP
                ├─ Obj_ind: NP — *Jo*
                ├─ Obj_dir: NP — *two*
                └─ Mod: PP — *after*

5.3.2 Gapped coordination

CGEL calls examples like *Kim is$_x$ an engineer and Pat __$_x$ a barrister* "gapped coordinations" and includes gaps in the phrase structure. In contrast, CGELBank treats them mostly like right nonce-constituent coordinations:

(64)

Coordination
├─ Coordinate: Clause
│   ├─ Subj: NP — *Kim*
│   └─ Head: VP — *is an engineer*
└─ Coordinate: NP + NP
    ├─ Marker: Coordinator — *and*
    └─ Head: NP + NP
        ├─ Subj: NP — *Pat*
        └─ PredComp: NP — *a barrister*

This also applies to more complex cases in which the "gap" is not a constituent but a string like *wanted him to marry*

(65)

Coordination

    Coordinate:
    Clause

    Coordinate:
    NP + NP

Subj:
NP

Head:
VP

Marker:
Coordinator

Head:
NP + NP

Subj:
NP

Obj:
NP

his father   *wanted him to marry sue*   *but*   *his mother*   *Louise*

An interesting case occurs in coordinated verbless complements of *with*, as in (66). This example also illustrates that in gapped coordination the function of the coordination, if it is not the root, will usually be a typical function like Comp, as opposed to the kind of nonce function found in right nonce-constituent coordination.

(66)

PP

Head:
P

Comp:
Coordination

Coordinate:
NP + AdjP

Coordinate:
NP + PP

Subj:
NP

PredComp:
AdjP

Marker:
Coordinator

Head:
NP + PP

Subj:
NP

Comp:
NP

with   *Jill*   *intent on staying*   *and*   *Pat*   *on leaving*

The coordinates in this construction can have more than two constituents, as in (67), which is the same string as (63) but semantically and structurally quite different.

(67)

Coordination

Coordinate:
Clause

Coordinate:
NP + NP + PP

Subj:
NP

Head:
VP

Marker:
Coordinator

Head:
NP + NP + PP

Head:
VP

Mod:
PP

Subj:
NP

Obj$_{dir}$:
NP

Mod:
PP

Head:
V

Obj$_{ind}$:
NP

Obj$_{dir}$:
NP

Mo   *gave*   *me*   *one*   *before*   *and*   *Jo*   *two*   *after*

5.3.3 Delayed right constituent coordination

CGELBank treat these constructions as having a gap and a postnucleus.

(68)

```
                                    Clause
                   _____|_____
                  |                                       |
               Subj:                                    Head:
                NP                                        VP
                                         _____|_____
                                        |                                     |
                                      Head:                               Postnucleus:
                                   Coordination                               NPₓ
                            _____|_____
                           |                     |
                      Coordinate:           Coordinate:
                          VP                    VP
                      ____|____            _____|_____
                     |         |          |           |
                   Head:     Obj:      Marker:      Head:
                    V        GAPₓ    Coordinator     VP
                                                 _____|_____
                                                |           |
                                              Head:        Comp:
                                              Vaux        Clause
                                                            |
                                                          Head:
                                                           VP
                                                       _____|_____
                                                      |           |
                                                    Head:        Obj:
                                                     V          GAPₓ

     I     saw      –          but    didn't   meet      –     her
```

CGEL takes no clear position on cases like *He's as old as or older than me*, where there is no prosodic break before the final element, and where it can be an unstressed personal pronoun (p. 1345). CGELBank takes the same approach as above, as illustrated in (69).

(69)

AdjP

Tree diagram (69):

- AdjP
  - Head: Coordination
    - Coordinate: AdjP
      - Head: AdjP — *as old*
      - Comp$_{ind}$: PP
        - Head: P — *as*
        - Comp: GAP$_x$ — –
    - Coordinate: AdjP
      - Marker: Coordinator — *or*
      - Head: AdjP
        - Head: Adj — *older*
        - Comp: PP
          - Head: P — *than*
          - Comp: GAP$_x$ — –
  - Postnucleus: NP$_x$ — *me*

5.3.4   End-attachment coordination

Postposing of coordinate

(70)

Tree diagram (70):

- Clause
  - Subj: NP — *I*
  - Head: VP
    - Head: VP
      - Head: V — *made*
      - Obj: Coordination
        - Coordinate: NP — *this one*
        - Coordinate: GAP$_x$ — –
      - PredComp: AdjP — *too sweet*
    - Postnucleus: NP$_x$ — *but not that one*

Addition of a new element

This construction is not, in fact, a coordination. Instead, the coordinator-initial constituent that looks like a final coordinate is a supplement (see also §3.1.3 and §5.4.5).

(71)

```
                    Clause
          ┌───────────┬───────────────┐
        Subj:       Head:         Supplement:
         NP          VP              AdvP
          △          ╱╲              ╱╲
          I       knew her        but not well
```
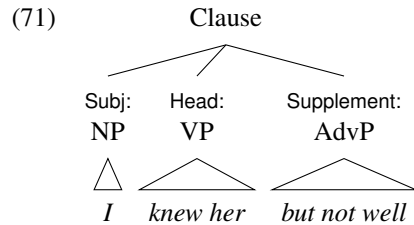
As this is not a coordinate, however, the Coordinator is allowed to be sister to a lexical head, subject to the usual preference for binary branching:

(72)

```
                    Clause
          ┌───────────┬───────────────────┐
        Subj:       Head:            Supplement:
         NP          VP                 AdvP
          │           │           ┌───────┴───────┐
          │           │        Marker:          Head:
          │           │      Coordinator         Adv
          │           │           │               │
          I        met her       but           briefly
```

5.4 Individual constructions with coordinators

5.4.1 *Not only* X *but* Y

At the beginning of a *but*-coordination, *not* (*only/just/even…*) bears parallels to a marker of correlative coordination as in §5.2.5 (*either… or*, *both… and*, etc.), but is better analyzed as a modifier (pp. 1313–1314).

(73)

Coordination

```
                    Coordination
          ┌──────────────┴──────────────┐
    Coordinate:                    Coordinate:
      Clause                         Clause
   ┌─────┴─────┐              ┌─────────┴─────────┐
  Subj:      Head:        Marker:              Head:
   NP         VP        Coordinator           Clause
   │      ┌────┴────┐        │            ┌──────┴──────┐
   │    Head:    PredComp:   │          Subj:        Head:
   │    V_aux      AdjP      │           NP            VP
   │      │     ┌────┴────┐  │           │        ┌─────┴─────┐
   │      │    Mod:    Head: │           │      Head:     PredComp:
   │      │    AdvP    Adj   │           │      V_aux       AdjP
   │      │      │       │   │           │        │           │
   he    was  not only right but         he      was      prescient
```
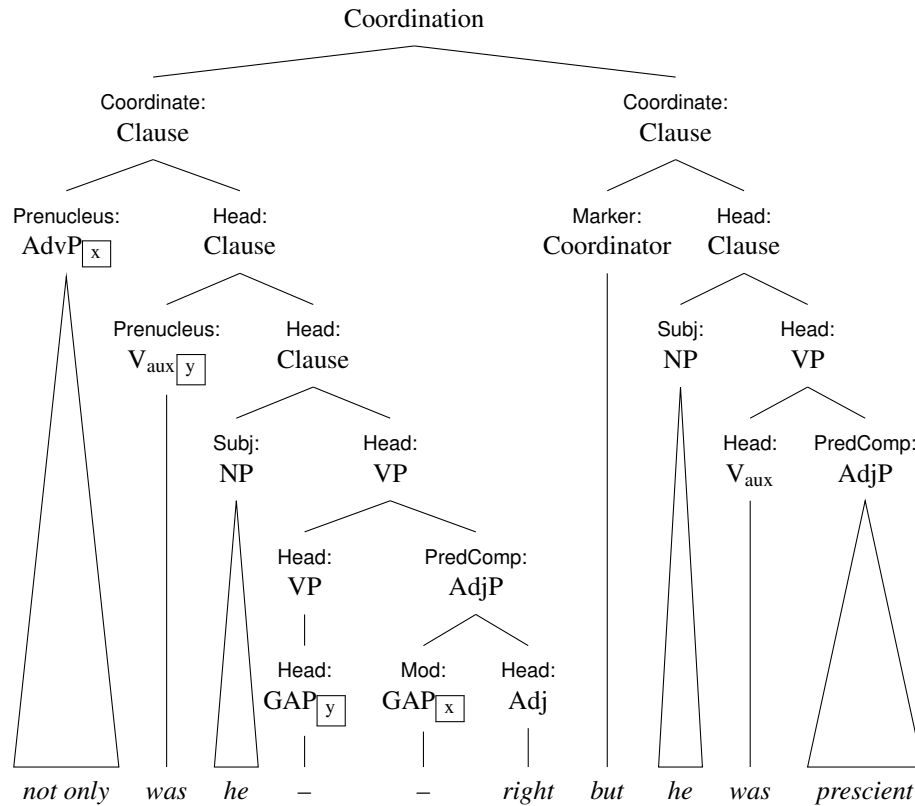
The construction exemplified in (74) is a special case of inversion (see §4.2.3):

56

(74)

Coordination

Coordinate: Clause — Coordinate: Clause

Prenucleus: AdvP $_x$ — Head: Clause

Marker: Coordinator — Head: Clause

Prenucleus: V$_{aux}$ $_y$ — Head: Clause

Subj: NP — Head: VP

Subj: NP — Head: VP

Head: VP — PredComp: AdjP

Head: V$_{aux}$ — PredComp: AdjP

Head: VP — PredComp: AdjP

Head: GAP $_y$ — Mod: GAP $_x$ — Head: Adj

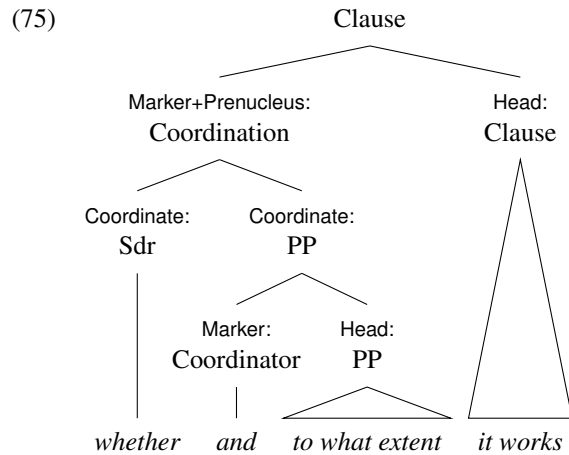*not only*   *was*   *he*   –   –   *right*   *but*   *he*   *was*   *prescient*

If *but* were omitted from the sentence, it would be analyzed as asyndetic coordination (§5.2.1, but see CGEL p. 1314 noting that this analysis is debatable).
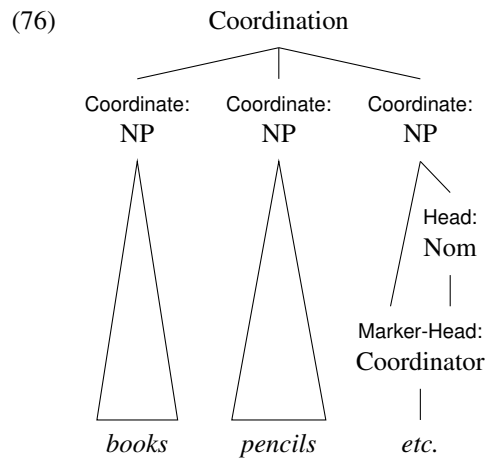
5.4.2   *Whether*

Whether is an unusual subordinator in what it coordinates with. The expressions *whether or not* and *whether or no* are complex subordinators (see §2.3.2) and not coordinations. Nevertheless, *whether* can coordinate with interrogative phrases (e.g., *whether and why we would go*), including PPs as in *whether and to what extent*. Examples like those in (75) are the only cases where a coordinate is a lexeme.

(75)

Clause

Marker+Prenucleus:
Coordination

Head:
Clause

Coordinate:
Sdr

Coordinate:
PP

Marker:
Coordinator

Head:
PP

whether   *and*   *to what extent*   *it works*

5.4.3 *Etc.*

Etc. is literally "and others", so we take it to be a coordinator in fused marker-head function (see §3.2).

(76)

Coordination

Coordinate:
NP

Coordinate:
NP

Coordinate:
NP

Head:
Nom

Marker-Head:
Coordinator

books   *pencils*   *etc.*

5.4.4 x to y ranges

We take cases like *5:00–8:00* in phrases like *between 5:00–8:00* or on their own to be coordinations, with the hyphen - or – or being a coordinator read "and" or "to" or having no phonological realization.[1] The slash */* is a possible alternative which can also mean "or" (p. 1764). There are also instances with just a space, such as *this Earth Moon highway*. We take these to be asyndetic coordinations.
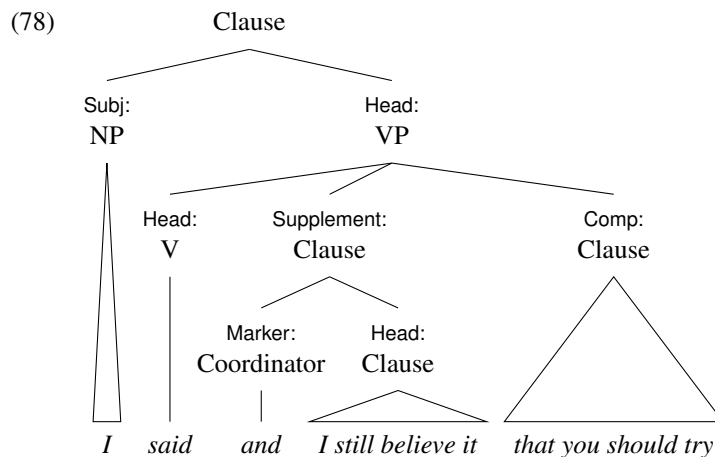
---

[1] These should not be confused with coordinative compounds such as *Austria-Hungary*, *Hewlett-Pakard*, or *murder-suicide*, which, like all compounds, are individual lexical items.

(77) a. *I work <u>Mon–Fri</u>/<u>Mon to Fri</u>.*
     b. *John Nash (<u>1928–2015</u>)*
     c. *a <u>French–English</u>/<u>French/English</u> dictionary*
     d. *the <u>May/June</u> period*
     e. *a <u>parent–teacher</u> meeting*
     f. *pp. <u>23–64</u>*

In contrast, in constructions like *We went from Toronto to Georgetown*, the PPs *from Toronto* and *to Georgetown* are a source complement and a goal complement respectively (p. 258). And the structure of cases like <u>*From Toronto to Georgetown* /</u> <u>*To Georgetown from Toronto is about eight and a half hours*</u> has the second PP as a complement in the first (p. 433).

5.4.5   Supplements marked by a coordinator

CGEL notes that supplements may be marked by a coordinator (p. 1361).[2] An example showing the representation in CGELBank is given in (78). See §3.1.3 for the structure of supplements.

(78)

---

[2]A sentence consisting solely of a constituent marked by a coordinator is not a supplement (see §5.2.3).

59

Chapter 6

Miscellaneous: Ellipsis, errors, punctuation, and future goals

6.1 Ellipsis

As noted in §4.7.3, we do not use "gaps" to show ellipted material. The following examples are provided to illustrate where a gap would **not** appear in CGELBank.

6.1.1 Elliptical stranding (Post-auxiliary ellipsis, including subordinator *to*)

CGEL sometimes marks gaps in examples with post-auxiliary ellipsis such as *I'll help you$_x$ if I can . . . $_x$.* In such a case, the contents of the gap may be inferred from elements in a higher clause, but it might have to be inferred from the context (e.g., *Do you think we should . . . ?*). This means it's not always possible to have a co-index.

6.1.2 Ellipsis of postmodifiers

[*An article on this topic$_x$*] *is more likely to be accepted than* [*a book . . . $_x$*].

6.1.3 In response to questions

A: *Whose father$_x$ is on duty today?* B: *Kim's . . . $_x$.*

6.1.4 With *let's*

A: *Let's go$_x$.* B: *Yes, let's . . . $_x$.*

6.1.5 Reduced interrogative clauses

He made some mistakes$_x$, though I don't know how many . . . $_x$.

6.1.6 Subclausal coordination

There is a copy$_x$ [*on the desk and . . . $_x$ in the top drawer*].
We'll be in$_x$ Paris for a week and . . . $_x$ Bonn for three days. (see Ch. 5 and CGEL's fn 65 on p. 1343)

6.1.7 Ellipsis of complement of lexical verbs and adjectives

A: *Hawaii$_x$ would be nice, but I can't go . . . $_{xy}$.* B: *But you promised . . . $_y$.*

6.1.8 Ellipsis of subject

In CGELBank, such sentences are treated as VPs.

Ellipsis of personal pronoun subject

. . . Doesn't matter. (*it*)

Ellipsis of subject pronoun + auxiliary

...*Never seen anything like it!* (*I have*)

Ellipsis in closed interrogatives of auxiliary or auxiliary + subject pronoun

...*Feeling any better?* (*are you*)

6.1.9   Determiner in NP structure

...*Trouble is, we have to be there by six.* (*the*)

6.1.10   Radical ellipsis in open interrogatives

Such sentences are usually phrases. For example, the following is an AdvP.
    A: *I'm leaving.* B: *Why ....* (*are you leaving*)

6.1.11   Subordinate clauses

Cases like this are treated as phrases. For example, the complement of *wonder* in the following is an AdvP.
    A: *They got in without a key$_x$.* B: *I wonder how ...$_x$.*

6.1.12   Radical ellipsis in declarative responses

Again, such sentences are phrases. Here, *yesterday* is an NP.
    A: *When did she get home?* B: *... yesterday.* (*she got home*)

6.2   Errors

Where there is a clear error, both the original form and the corrected form are included in the tree: the original form with a :t value (for "token" or "terminal"), and the corrected form with a :correct value. The lemma should reflect the correct version of the word, and must be indicated explicitly if it differs from the :correct value.

In (79), for instance, the **misspelling** *out* has been corrected to *our*, and its lemma is *we*. In the graphical display, the corrected form will be included in parentheses after the original form.[1]

(79) (N_pro :t "out" :correct "our" :l "we")

---

[1]The overarching description of the raw data format appears in Appendix B.

6.2.1 Omissions

Consider *They are preparing my older son for kindergarten and looks forward to seeing his teacher and friends everyday.* Presumably, it is the son looking forward to things, so the subject of *looks* should probably be *he*. A pronoun token is thus inserted for the missing word; it has no original token (`:t` value), so it is possible to recognize it as an insertion.

(80) `(N_pro :correct "he")`

6.2.2 Extra words

The example *go to the room 401* illustrates an apparent grammatical error: *the* should be omitted in this context, though it is close to its ordinary use as a determinative in determiner function. We retain the word in the tree with an empty string value for `:correct`:

(81) `(D :t "the" :correct "")`

We have not yet encountered any sentences with extra words due to speech repair or total incoherence. Such words might be deemed unparseable and removed from the tree.

6.2.3 Non-standard morphology

Where this seems to be a dialectal form, no correction is indicated. Where it appears to be an error, it's treated as a misspelling. For example *I am works hard* should presumably be *I am working hard*.

6.2.4 Punctuation

Punctuation is considered to have little impact on tree structure (see §6.3) and is not included in the graphical trees. In the raw trees, punctuation tokens are included as `:p` values alongside lexical tokens:

(82) `(Adj :p "?" :p ")" :p "," :t "heartless")`

Note that there can be multiple `:p` tokens within a lexical node, and they may precede and/or follow the lexical token as reflects the order in the sentence. The current convention does not place any weight on the semantics of the punctuation (such as the difference between open-parentheses/quotes and close-parentheses/quotes), but uniformly places punctuation tokens before the next lexical token if there is one, as in (82). At the end the sentence, any punctuation tokens go after the last lexical token.

CGELBank does not include any corrections for nonstandard use or omission of punctuation tokens (i.e., punctuation marks that fall outside of words).

6.2.5 Syntax

Apart from errors noted above, CGELBank does not annotate errors of syntax. For example, the verb *said* does not license an indirect object. Nevertheless, in *I said him the answer*, no correction would be noted.

6.3 Punctuation and symbols

In CGELBank, punctuation is included in raw trees (as described in §6.2.4) but omitted from graphical trees. Syntactically, punctuation often has little impact except for marking the end of a sentence or marking supplements.

6.3.1 Primary terminals

A tree typically consists of a sentence, which ends with a primary terminal.

6.3.2 Marking supplements

Supplements are often set off by punctuation, as in (83).

(83) a. *An official, who spoke on the condition of anonymity, said it was unlikely to succeed.*
      b. *Anderson and Der Khosla of the Bureau of Competition Policy (the "Bureau") defined industrial policy as . . .*
      c. *It was questionable at the time; however, it's now the consensus view.*

6.3.3 Hyphens

See §5.4.4.

6.3.4 Position of currency symbols

Currency symbols like *$* for *dollar(s)* should be placed in the order pronounced, regardless of the original orthography. That is, the untokenized string *$300* would be tokenized and ordered in the tree as *300 $*, and analyzed with the same structure as *300 dollars*. This is necessary as the quantity phrase may be complex (e.g., *over $300* analyzed as *[over 300] dollars*).

6.3.5 Emoji

We treat emoji as interjections in supplement function.

6.4 Future goals

As additional data is encountered, additional clarifications to the guidelines are expected to become necessary.

Based on the data encountered already, we feel that a few areas merit further deliberation or enhancement.

6.4.1 Potential points of revision

Verbless clauses.   The structure of what CGEL terms verbless clauses (e.g., *With the baby asleep, we can go about our business*) is unclear. Is it appropriate to consider them a subtype of clause if there is no verb? Further discussion is necessary.

External modifiers.   Within NPs, CGEL describes a distinction between internal and external modifiers. The external ones are outside the Nom, and thus structurally identifiable in NPs. But some of them—notably focusing modifiers—can also modify other kinds of phrases. Is a designated external modifier function warranted?

Medial modifiers.   CGEL speaks of a notion of *core* complement in a VP. Objects, perhaps, are core whereas PP complements are not, as evidenced by adverbs' resistance to placement between verb and object. But it is not clear how uniform this pattern is. Does it warrant a revision to the branching rules within VPs?

Richer treatment of structure within special name patterns, dates, etc.   CGEL is largely silent on this point.

6.4.2 Potential enhancements

On the whole, the current granularity of category and function labels has proved workable. We are reluctant to make them finer-grained as this would make it more difficult to create and interpret the main tree structures.

However, additional features or forms of annotation could be added to supplement what is already in the tree.

Morphology.   While parallel UD parses provide morphological information on tokens, the terminology differs somewhat from that of CGEL. Morphosyntactic features could be added at the phrase level as well.

Clause types.   Currently, the only explicit subtypes of clause in CGELBank are relative clauses. But CGEL defines an inventory of clause types (e.g., open interrogative, closed interrogative, exclamative) which might be added as supplementary information.

Ellipsis.   §6.1 indicates different kinds of ellipsis, but ellipsis is not made explicit in CGELBank at present.

Constructions beyond surface structure.   Formal frameworks often contain mechanisms for constructions such as passives, raising, and control. Such constructions are, of course, addressed in CGEL, but not made explicit in trees. Perhaps they should be added as an additional layer somehow. The licensors of indirect complements (§3.3) and the distinction between predicative and non-predicative adjuncts could be indicated as well.

Acknowledgements

Bibliography

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, July 2021. URL https://doi.org/10.1162/coli_a_00402.

Rodney Huddleston and Geoffrey K. Pullum, editors. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK, 2002. URL https://archive.org/details/TheCambridgeGrammarOfTheEnglishLanguage.

Rodney Huddleston, Geoffrey K. Pullum, and Brett Reynolds. *A Student's Introduction to English Grammar*. Cambridge University Press, 2nd edition, 2021. DOI: 10.1017/9781009085748.

Christian M. I. M. Matthiessen and John A. Bateman. *Text Generation and Systemic-functional Linguistics: Experiences from English and Japanese*. Pinter, 1991.

John Payne, Rodney Huddleston, and Geoffrey K. Pullum. Fusion of functions: The syntax of *once*, *twice* and *thrice*. *Journal of Linguistics*, 43(3):565–603, November 2007. URL https://www.cambridge.org/core/journals/journal-of-linguistics/article/fusion-of-functions-the-syntax-of-once-twice-and-thrice1/D0D583318480CCEB18536C891286D48E. Cambridge University Press.

John Payne, Rodney Huddleston, and Geoffrey K. Pullum. The distribution and category status of adjectives and adverbs. *Word Structure*, 3(1):31–81, April 2010. URL https://www.euppublishing.com/doi/abs/10.3366/E1750124510000486. Edinburgh University Press.

John Payne, Geoffrey K. Pullum, Barbara C. Scholz, and Eva Berlage. Anaphoric *one* and its implications. *Language*, 89(4):794–829, 2013. URL https://www.jstor.org/stable/24671958. Linguistic Society of America.

Geoffrey K. Pullum and Brett Reynolds. New members of 'closed classes' in English. Manuscript, February 2013. URL https://www.researchgate.net/publication/260122411_New_members_of_%27closed_classes%27_in_English.

Geoffrey K. Pullum and James Rogers. Expressive power of the syntactic theory implicit in *The Cambridge Grammar of the English Language*. In *Annual Meeting of the Linguistics Association of Great Britain*, pages 1–16, 2009. URL http://www.lel.ed.ac.uk/~gpullum/EssexLAGB.pdf.

Brett Reynolds, Aryaman Arora, and Nathan Schneider. Unified syntactic annotation of English in the CGEL framework. In *Proc. of the 17th Linguistic Annotation Workshop (LAW-XVII)*, Toronto, Canada, July 2023. URL https://people.cs.georgetown.edu/nschneid/p/cgeltrees.pdf.

Nathan Schneider and Amir Zeldes. Mischievous nominal constructions in Universal Dependencies. In *Proc. of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021)*, pages 160–172, Sofia, Bulgaria, December 2021. URL https://aclanthology.org/2021.udw-1.14.

Appendix A

Justification of lexical coordination analysis

This appendix offers an in-depth justification of the analysis of lexical coordination presented in §5.2.2.

Coordination of lexemes presents the puzzle of what category to assign the expanded coordinate. Strictly, *and will* in (84–88) is neither a verb, a kind of lexeme, since it has internal structure; nor is it a typical VP as it will not admit of any complement. A number of possible analyses present themselves.

1. Call it a V and ignore the internal structure in a lexical category as in (84).

2. Forbid coordination of lexemes, and introduce a new set of marked categories which are atypical phrases as in (85).

3. Call all non-marked single coordinates Vs and call marked coordinates VPs, as in (86), accepting that such a VP will not allow a complement. (This fact can actually be motivated by the preference for binary branching, noting that a VP like *easily do it* also does not allow a complement.)

4. ☞ Treat auxiliaries in coordination as projecting unary VPs as the coordinates (87).

5. Forbid coordination of lexemes, call it a VP, and include a gap co-indexed to the post-nuclear complement as in (88). This makes it a delayed right constituent coordination (see §5.3.3).

The last option seems appealing for simple cases like (88) and (89), but for cases like (90), it would require the head of the main clause to be co-indexed to a prenucleus (because of Subj-Aux inversion), to have that prenucleus have coordinated phrases with gaps and a post-nucleus, but the post-nucleus would not be realized there. Instead it would be a gap co-indexed to a post-nucleus in the main VP. This seems beyond the pale.

On balance, we feel option 4 (unary VP coordinates) offers the best compromise. It preserves coordination of like categories, does not require a new category for marked coordinates, and does not apply a lexical category to marked coordinates, which are phrases. Finally, it adheres to the principle that all lexemes other than subordinators and coordinators project a phrasal category (§2.2.1).

69

(84) V$_{aux}$ marked coordinate

```
                    Clause
         _____/      _____
     Subj:                          Head:
      NP                             VP
       |              _____/   _____
       |          Head:                     Comp:
       |       Coordination                 Clause
       |        ___/    \___                    |
       |   Coordinate:   Coordinate:            |
       |    ┌──────┐      ┌──────┐              |
       |    │ V_aux│      │ V_aux│              |
       |    └──────┘      └──────┘              |
       |       |          __/   \__             |
       |       |      Marker:     Head:         |
       |       |    Coordinator  ┌──────┐       |
       |       |        |        │ V_aux│       |
       |       |        |        └──────┘       |
       I       can      and       will        try
```

(85) V$_{aux}$^{mk}

```
                    Clause
         _____/      _____
     Subj:                          Head:
      NP                             VP
       |              _____/   _____
       |          Head:                     Comp:
       |       Coordination                 Clause
       |        ___/    \___                    |
       |   Coordinate:   Coordinate:            |
       |    ┌──────┐      ┌──────────┐          |
       |    │ V_aux│      │ V_aux^mk │          |
       |    └──────┘      └──────────┘          |
       |       |          __/   \__             |
       |       |      Marker:     Head:         |
       |       |    Coordinator  ┌──────┐       |
       |       |        |        │ V_aux│       |
       |       |        |        └──────┘       |
       I       can      and       will        try
```

(86) VP marked coordinate

```
                    Clause
         _____/      _____
     Subj:                          Head:
      NP                             VP
       |              _____/   _____
       |          Head:                     Comp:
       |       Coordination                 Clause
       |        ___/    \___                    |
       |   Coordinate:   Coordinate:            |
       |    ┌──────┐      ┌────┐               |
       |    │ V_aux│      │ VP │               |
       |    └──────┘      └────┘               |
       |       |          __/   \__             |
       |       |      Marker:     Head:         |
       |       |    Coordinator  ┌──────┐       |
       |       |        |        │ V_aux│       |
       |       |        |        └──────┘       |
       I       can      and       will        try
```
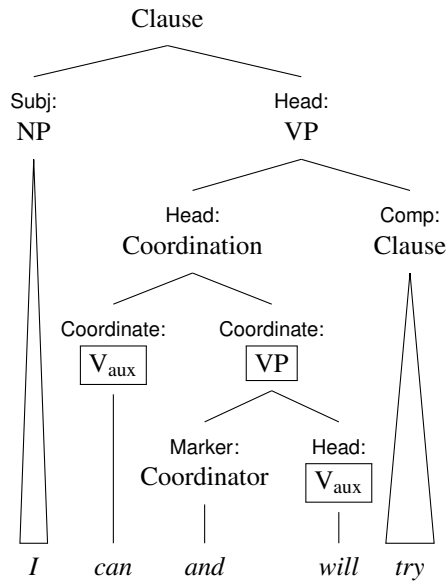
(87) Unary VP coordinates

```
                      VP
          _____/    _____
      Head:                        Comp:
   Coordination                    Clause
    ___/    \___                      |
Coordinate:   Coordinate:             |
 ┌────┐        ┌────┐                 |
 │ VP │        │ VP │                 |
 └────┘        └────┘                 |
    |          __/   \__              |
 Head:     Marker:     Head:          |
 V_aux   Coordinator  ┌────┐          |
    |        |        │ VP │          |
    |        |        └────┘          |
    |        |        Head:           |
    |        |        V_aux           |
   can       and       will         try
```

(88)   Gapped VP coordinates

Clause

Subj:
NP

Head:
VP

Head:
Coordination

Postnucleus:
Clause$_x$

Coordinate:
VP

Coordinate:
VP

Head:
V$_{aux}$

Comp:
GAP$_x$

Marker:
Coordinator

Head:
VP

Head:
V$_{aux}$

Comp:
GAP$_x$

I   *can*   *–*   *and*   *will*   *–*   *try*

(89)

Clause

- Prenucleus: NP_x
- Head: Clause
  - Prenucleus: $V_{aux\,y}$
  - Head: Clause
    - Subj: NP
    - Head: VP
      - Head: GAP_y
      - Comp: Clause
        - Head: VP
          - Head: Coordination
            - Coordinate: VP
              - Head: V
              - Obj: GAP_x
            - Coordinate: VP
              - Marker: Coordinator
              - Head: VP
                - Head: V
                - Obj: GAP_x
          - Postnucleus: NP_x

what *are* *you* – *eating* – *and* *drinking* – –

72

(90)

Clause

Prenucleus:
NP_x

Head:
Clause

Prenucleus:
VP_z

Head:
Clause

Head:
Coordination

Postnucleus:
NP_y

Subj:
NP

Head:
VP

Coordinate:
VP

Coordinate:
VP

Head:
GAP_z

Postnucleus:
$Clause_y$

Head:
V_{aux}

Comp:
GAP_y

Marker:
Coordinator

Head:
VP

Head:
VP

Head:
V_{aux}

Comp:
GAP_y

Head:
V

Obj:
GAP_x

what *did* – *or* *will* – – *you* – *decide* –

73

Appendix B

Data Format

The data release can be accessed at https://github.com/nert-nlp/cgel. The raw data is stored in .cgel files, one per subcorpus: currently, ewt.cgel for the EWT trees and twitter.cgel for the Twitter trees. These files can be opened in a text editor. A Python API for loading the trees is provided for scripting (see cgel.py), and there is also a tool (tree2tex.py) to generate LaTeX to produce graphical versions of the trees.

An example raw tree is given in Figure B.1. Each tree has two parts: the header/ metadata section consisting of lines beginning with #, and the tree itself.

Lines in the header are key-value pairs following the .conllu standard used in the Universal Dependencies project. In particular, every sentence has an ID as well as a number indicating its order within the file. The text line indicates the original, untokenized sentence. The sent line is the sequence of terminals to appear in the CGEL tree: punctuation tokens are removed, capitalization is normalized, and -- is inserted at positions where there are gaps.

The tree itself is in a parenthesized format adapted from PENMAN notation (Matthiessen and Bateman, 1991). Every line represents the start of a constituent. Parentheses (and accompanying indentation) indicate the bracketing structure. Functions begin with a : symbol and are capitalized. Gaps in the tree consist of the GAP category. A coindexation variable and slash precede gaps and their coindexed constituents.

B.1 Features

Table B.1 lists the features that may be indicated within a node following its category. These principally apply to lexical nodes, but a :note can appear on any node. String values for these features must be delimited by double quotes. Two escapes are provided: \" for the literal quotation mark character and \\ for the backslash character.

B.2 Fusion

Fusion of functions, which occurs in 2 places in Figure B.2, is not expressed directly in the raw tree format. The parenthesized notation pretends that the first of the two

```
# sent_id = Tree IsThatWhatYouCall-0
# sent_num = 4
# text = Is that what you call WH-movement?
# sent = is that -- what you call -- WH-movement
(Clause
    :Prenucleus (x / VP
        :Head (V_aux :t "is" :l "be"))
    :Head (Clause
        :Subj (NP
            :Head (Nom
                :Det-Head (DP
                    :Head (D :t "that"))))
        :Head (VP
            :Head (x / GAP)
            :PredComp (NP
                :Head (Nom
                    :Mod (Clause_rel
                        :Head-Prenucleus (y / NP
                            :Head (Nom
                                :Head (N_pro :t "what")))
                        :Head (Clause_rel
                            :Subj (NP
                                :Head (Nom
                                    :Head (N_pro :t "you")))
                            :Head (VP
                                :Head (V :t "call")
                                :Obj_dir (y / GAP)
                                :Obj_ind (NP
                                    :Head (Nom
                                        :Head (N :t "WH-movement"
                                            :subt "WH" :subt "-"
                                            :subt "movement" :p "?")))))))))))))
```

Figure B.1: Example raw tree from twitter.cgel (with extra line breaks in the final *WH-movement* constituent so it doesn't overflow the margin). The graphical view of this tree is in Figure B.2.

:note	a comment on the analysis
:p	punctuation token before or after a word token (may be used multiple times in the node; see §6.2.4)
:t	token value: the original form of the word after tokenization (leaf nodes only; GAPs and words inserted to correct an omission have no :t)
:subt	subtoken (used multiple times per node) for words where the Universal Dependencies tokenization is finer-grained, e.g. possessive clitics
:correct	corrected form (see §6.2)
:l	lemma *when distinct from the word form*

Table B.1: String-valued features that may be specified on nodes in the .cgel format.

Clause

Prenucleus:
V_{auxx}

Head:
Clause

Subj:
NP

Head:
VP

Head:Head:
NomGAP$_x$

PredComp:
NP

Det-Head:
DP

Head:
Nom

Head:
D

Mod:
Clause$_{rel}$

Head-Prenucleus:
NP$_y$

Head:
Clause$_{rel}$

Head:
Nom

Subj:
NP

Head:
VP

Head:
N$_{pro}$

Head:
Nom

Head:
V

Obj$_{dir}$:
GAP$_y$

Obj$_{ind}$:
NP

Head:
N$_{pro}$

Head:
Nom

Head:
N

is   that   –   what   you   call   –   WH-movement
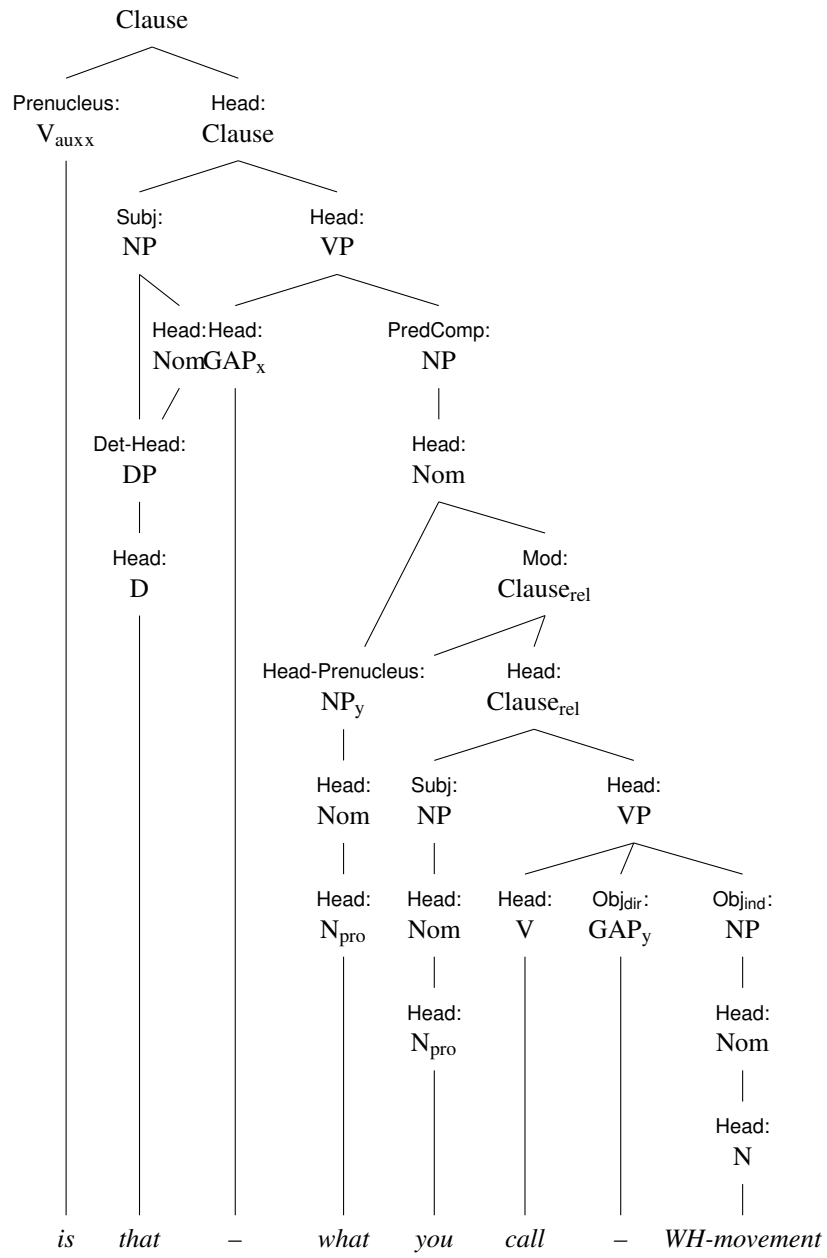
Figure B.2: Graphical view of the tree from Figure B.1.

incoming branches (the one skipping a level) is missing, as in Figure B.1. The branch can be recovered automatically by attaching each constituent in Det-Head, Mod-Head, Marker-Head, or Head-Prenucleus function to its non-immediate ancestor (typically grandparent) of the appropriate category (this extra attachment reflects the first part of the hyphenated function). For example, the deeper parent of a constituent labeled Det-Head will be a Nom, and its additional parent will be the NP headed by that Nom (or, if there is layering as in (19), another Nom which it heads).