Bracketing Guidelines for Treebank II Style Penn Treebank Project [1]

Principal authors:
Ann Bies, Mark Ferguson, Karen Katz, and Robert MacIntyre
Major contributors:
Victoria Tredinnick, Grace Kim, Mary Ann Marcinkiewicz, Britta Schasberger [2]

January 1995

[1]This phase of the project was funded by the Linguistic Data Consortium. Previous work was funded by DARPA and AFOSR jointly under grant No. AFOSR-90-006, with additional support by DARPA grant No. N0014-85-K0018 and by ARO grant No. DAAL 03-89-C0031 PRI. Seed money was provided by the General Electric Corporation under grant No. J01746000. We gratefully acknowledge this support.

[2]We would like to thank Mitch Marcus for his support and encouragement in the production of this document and the policy it describes. Leslie Dossey and Elizabeth Hamilton put a lot of effort into early analysis and organization of the issues. Beatrice Santorini wrote the previous manual, upon which much of our policy is still based. Finally, we would like to thank a set of people too numerous to mention specifically for their helpful criticisms, suggestions, and advice.

Contents

A note about the examples in this manual:
Many of the examples in this manual are taken directly from the bracketed corpus, with full structure intact (excluding Part-of-Speech tags). Occasionally, however, we simplify the example by removing irrelevant internal structure; we hope that these occasions will be clear.

In addition, we usually omit final punctuation, as well as the outermost unlabeled parentheses that surround top-level constituents in the actual data files.

1 An Overview of Basic Clause Structure

This section presents of an overview of basic structure, but it does not attempt to summarize or justify the entire policy. For an overview of the new bracketing style, with some notes about its usefulness, see [Marcus et al. 1994] (included on the CDROM release as "arpa94"). For an overview of the Treebank Project in general, see [Marcus et al. 1993] ("cl93"). Note also that, for the most part, this manual focuses on *problematic* constructions rather than *common* ones, for mostly historical reasons.

The basic structure of an S in the Treebank grammar was formerly (Preliminary Release, Version 0.5, 1992):

```
(S (NP Casey)
   (VP throws
       (NP the ball)))
```

More complicated structures were annotated with an AUX node.

```
(S (NP Casey)
   (AUX will)
   (VP throw
       (NP the ball)))


(S (NP Casey)
   (AUX should)
   (VP have
       (VP thrown
           (NP the ball))))
```

The basic approach to simple sentences has not changed. However, we no longer have a specially bracketed and labeled AUX; the node that was formerly AUX now corresponds to the highest level of the VP.

```
(S (NP-SBJ Casey)
   (VP will
       (VP throw
           (NP the ball))))


(S (NP-SBJ Casey)
   (VP should
       (VP have
           (VP thrown
               (NP the ball)))))
```

1.1 Basic elements of S

1.1.1 Predicate

The predicate is either the lowest (right-most branching) VP or (after copular verbs and in "small clauses") a constituent tagged -PRD. Moved predicates leave a coindexed trace *T* in VP.

1.1.2 Arguments of the predicate

- External:

  The **surface subject** is tagged -SBJ (subject).

- Internal:

  1. **Direct object NP**: occurs after the verb, has no function tag, and is not followed by another NP.
  2. **Indirect object NPs** are of the following types:

     (i) NPs that occur between the verb and its direct object and have no function tags, as in dative shift constructions (e.g., *gave* **Mary** *the book*)

     (ii) dative PPs (e.g., *to Mary*), tagged -DTV. Only verbs that can undergo dative shift are considered to have dative objects.

1.1.3 Level of attachment

- S-level:

  The following are attached at S-level: subject NP, highest VP, fronted constituents, initial and final punctuation, and most modifiers that precede the verb phrase. When there is no VP (as in "small clauses"), the predicate is labeled -PRD , and it and any following adjuncts are attached at S-level.

- VP-level:

  1. Almost all modifiers that follow the verb are attached under the lowest appropriate VP. When there is conjunction and the modifier applies to both VPs, the modifier is attached at conjunction level.
  2. An exception is made for modifiers that are interpreted as appositives to the event or the predicate itself. Such modifiers are adjoined to VP. Some of them may also have a -ADV tag.

```
(S (NP-SBJ Investors)
   (VP might
       (VP (VP appear
                (ADJP-PRD unenthusiastic
                          (PP about
                              (NP the new issue))))
           (SBAR (WHNP-1 which)
                 (S (NP-SBJ *T*-1)
                    (VP might
                        (VP force
                            (S (NP-SBJ the government)
                               (VP to
                                   (VP raise
                                       (NP the coupon)
                                       (PP-CLR to
                                               (NP (QP more than 7)
                                                   %))))))))))))


(S (NP-SBJ they)
   (VP would
       (VP (VP negotiate
                (NP rates)
```

```
                     (ADVP-MNR individually)
                     (PP-CLR with
                             (NP advertisers)))
                 ,
                 (NP-ADV (NP a practice)
                         (ADJP common
                               (PP-LOC in
                                       (NP broadcasting)))))))
```

```
(S (NP-SBJ factory inventories)
   (VP (VP fell
           (NP-EXT 0.1 %)
           (PP-TMP in
                   (NP September)))
       ,
       (NP (NP the first decline)
           (PP-TMP since
                   (NP February 1987)))))
```

1.1.4 Complementation within syntactic categories

The complement is attached inside the VP, NP, ADJP, or PP.

- Verbs:

  The term "complement" as it is used here refers to:

  1. internal arguments such as NP objects, S and SBAR with no adverbial dash tags (including some *if*-clauses, as in *I wonder if the Cubs are winning*), and quoted constituents (including SINV and FRAG)
  2. the passive logical-subject *by*-phrase
  3. VP
  4. constituents tagged -BNF, -CLR, -DTV, -PRD, and -PUT

```
(S (NP-SBJ-1 the guide)
   (VP was
       (VP given
           (NP *-1)
           (PP-DTV to
                   (NP Arthur))
           (PP by
               (NP-LGS Ford)))))
```

```
(S (NP-SBJ-1 Casey)
   (VP ought
       (S (NP-SBJ *-1)
          (VP to
              (VP have
                  (VP thrown
                      (NP the ball)))))))
```

- Nouns:

  Since it is difficult to consistently annotate an argument/adjunct distinction, all PP modifiers of nouns are Chomsky-adjoined to the NP:

```
(NP (NP a teacher)
    (PP of
        (NP chemistry)))
```

  However, clausal complements are recognized:

```
(NP the belief
    (SBAR that
          (S the world is flat)))
```

- Adjectives:

  Except in comparatives, any modifier following an adjective is bracketed as a complement.

```
(ADJP eager/likely/ready
      (S to believe anything))
```

```
(ADJP full
      (PP of
          (NP life)))
```

- Prepositions:

  The NP or S complement of a preposition is placed inside the PP.

1.1.5 Modification

- Premodifiers:

  Premodifiers generally are placed inside the phrase they are associated with:

```
(NP the red ball)

(ADJP extremely delicious)

(ADVP (NP one year) ago)
```

  VP premodifiers, however, are more often attached at S-level:

```
(S (NP-SBJ Sandy)
   (ADVP-TMP often)
   (VP throws
       (NP curves)))
```

  but they may also be attached inside the VP (see section 8 [Shared Complements and Modifiers] for more details):

```
(S (NP-SBJ Sandy)
   (VP (ADVP-MNR sneakily)
       threw
       (NP a curve)))
```

- Postmodifiers:

  Postmodifiers of NPs and comparative ADJPs are adjoined to the NP or ADJP:

```
(NP (NP a book)
    (PP about
        (NP toads)))
```

```
(ADJP (ADJP as tall)
      (PP as
          (NP him)))
```

  Postmodifiers of VP are attached under VP, with adverbial function tag(s) where appropriate:

```
(VP reading
    (PP-CLR about
            (NP toads))
    (PP-LOC on
            (NP the Internet)))
```

  When it is not clear whether a modifier within the VP should be attached at VP-level or to an object NP, the DEFAULT is to attach at VP-level (see section 5 [Pseudo-Attach]).

1.2 Clause types

We distinguish among a number of basic clause types: S, SINV, SBAR, RRC, SBARQ, SQ, S-CLF, *it*-extraposition, and FRAG.

1.2.1 S

- Simple declarative sentences:

```
(S (NP-SBJ Casey)
   (VP threw
       (NP the ball)))
```

- Passives:

  The surface subject is tagged -SBJ , the passive trace is indicated with (NP *) and coindexed to the surface subject, the *by*-phrase is a child of VP, and the logical subject is tagged -LGS . (Note that the -LGS tag goes on the NP and not on the PP of the *by*-phrase.)

```
(S (NP-SBJ-1 The ball)
   (VP was
       (VP thrown
           (NP *-1)
```

```
          (PP by
              (NP-LGS Casey)))))
```

- Imperatives:

  Imperatives are labeled S and given a null subject (NP-SBJ *).

```
(S (NP-SBJ *)
   (VP Throw
       (NP the ball))
   !)
```

If the name of the addressee appears with the imperative (at either the beginning or end), it is tagged -VOC (vocative) . The vocative is NOT coindexed to the null surface subject.

```
(S (NP-VOC American imperialists)
   ,
   (NP-SBJ *)
   (VP go
       (ADVP-DIR home))
   !)


(S (NP-SBJ *)
   (VP Close
       (NP the door)
       ,
       (NP-VOC John))
   .)
```

- Questions with declarative word order:

  Sentences that end with a question mark but have non-inverted word order are labeled S:

```
(S (NP-SBJ This)
   (VP is
       (NP-PRD Japan))
   ?)


(S (NP-SBJ You)
   (VP did
       (NP what))
   ?)
```

However, questions that are missing both subject and auxiliary are labeled SQ .

```
(SQ (NP-SBJ *)
    (VP See
        (NP that cute dog))
    ?)
```

- Infinitives:

  Infinitives are labeled S and take (NP-SBJ *) as the null subject, where *to* represents the highest level of the VP. (See section 14 [Infinitives] for more on the annotation of infinitivals.)

  1. Complement clauses.

     When the infinitive is a VP complement, the null subject of the infinitive is coindexed as usual to its logical subject (usually the subject of the matrix clause, but sometimes the object of the verb or not coindexed at all).

     ```
     (S (NP-SBJ-1 Casey)
        (VP wants
            (S (NP-SBJ *-1)
               (VP to
                   (VP throw
                       (NP the ball))))))
     ```

  2. Purpose clauses.

     Purpose clauses are attached at S and labeled -PRP (purpose/reason) . The subject is coindexed to the surface subject of the matrix clause when there is a coindexed interpretation.

     ```
     (S (NP-SBJ-1 Sue)
        (VP arrived
            (ADVP-TMP early)
            (S-PRP (NP-SBJ *-1)
                   (VP to
                       (VP get
                           (NP a good seat))))))
     ```

  3. Infinitival relatives.

     In the case of infinitival relatives, the relative is adjoined to NP and dominated by SBAR with a zero *wh*-complementizer labeled according to the role played by the gapped constituent. A *T* in the position of the gap is coindexed to the *wh*-complementizer. The (NP-SBJ *) is not indexed.

     ```
     (NP (NP a movie)
         (SBAR (WHNP-1 0)
               (S (NP-SBJ *)
                  (VP to
                      (VP see
                          (NP *T*-1))))))
     ```

- Participial and gerund clauses:

  Participial clauses have full clause structure, with either a lexical or null (NP-SBJ *) subject. (See section 13 [Gerunds and Participles] for more on the annotation of participial and gerund clauses.)

  "Floating" participles are tagged -ADV.

```
(S (S-ADV (NP-SBJ The crowd)
          (VP cheering
              (ADVP-MNR madly)))
   ,
   (NP-SBJ Willie)
   (VP caught
       (NP the ball)))
```

When appropriate, the null subject is coindexed to an NP in the matrix clause (generally the logical/structural subject).

```
(S (S-ADV (NP-SBJ *-1)
          (VP Running
              (PP-DIR toward
                      (NP Casey))))
   ,
   (NP-SBJ-1 Willie)
   (VP caught
       (NP the ball)))
```

Gerunds that act as the surface subject or as the object of a preposition are tagged -NOM .

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Baking
                  (NP pies)))
   (VP is
       (ADJP-PRD fun)))
```

```
(S (PP With
       (S-NOM (NP-SBJ interest rates)
              (VP rising)))
   ,
   (NP-SBJ the market)
   (VP is
       (VP moving
           (ADVP-MNR slowly))))
```

Verb complement gerunds , including those in “serial” verb constructions, are all bracketed as a simple S, with coindexed subject when appropriate.

```
(S (NP-SBJ I)
   (VP do not
       (VP mind
           (S (NP-SBJ you(r))
              (VP leaving
                  (ADVP-TMP early))))))
```

```
(S (NP-SBJ-1 The audience)
   (VP keeps
       (S (NP-SBJ *-1)
          (VP leaving
              (ADVP-TMP early)))))
```

SBAR complement gerunds are also bracketed as a simple S.

```
( (S (NP-SBJ-1 He)
     (VP ate
```

```
          (NP television)
          (SBAR-TMP while
                    (S (NP-SBJ *-1)
                       (VP watching
                           (NP dinner)))))
      .))
```

1.2.2 SINV

The SINV label is used for subject-auxiliary inversion in the case of negative inversion, conditional inversion, locative inversion, and some topicalizations. (SINV is not used with questions. See section 1.2.6 and section 1.2.5 for the treatment of subject-auxiliary inversion in the case of yes/no questions and *wh*-questions, respectively.) Inverted auxiliaries are unlabeled.

```
(SINV (ADVP-TMP Never)
      had
      (NP-SBJ I)
      (VP seen
          (NP such a place)))
```

When the inversion results in a conditional clause (i.e., when it is equivalent to (SBAR-ADV if...), the SINV is enclosed in SBAR-ADV).

```
(S (SBAR-ADV (SINV had
                   (NP-SBJ Casey)
                   (VP thrown
                       (NP the ball)
                       (ADVP-MNR harder))))
   ,
   (NP-SBJ it)
   (VP would
       (VP have
           (VP reached
               (NP home plate)
               (PP-TMP in
                       (NP time))))))
```

When subject-aux inversion is triggered by the predicate (or a piece thereof) moving out of the VP to a position preceding the subject, the moved predicate phrase is tagged -TPC (topicalized) and leaves a coindexed trace in the VP. Note that in this case, the auxiliaries that precede the subject are labeled VP (whereas in other cases of inversion they are left unlabeled), so that there's a place to properly attach the VP trace.

```
(SINV (VP-TPC-1 Marching
                (PP-DIR past
                        (NP the reviewing stand)))
      (VP were
          (VP *T*-1))
      (NP-SBJ 500 musicians))


(SINV (ADJP-PRD-TPC-2 (ADVP Even more)
```

```
                    unusual)
      (VP is
          (ADJP-PRD *T*-2))
      (NP-SBJ (NP the mating behavior)
              (PP of
                  (NP the praying mantis))))

(SINV (ADVP Also)
      (ADJP-PRD-TPC-11 present)
      (VP will
          (VP be
              (ADJP-PRD *T*-11)))
      (NP-SBJ (NP (NP the bride 's)
                  children)
              ,
              (NP Joan and Kirkland)))

(SINV (ADVP-DIR-TPC-3 Out)
      (VP might
          (VP have
              (VP popped
                  (ADVP-DIR *T*-3))))
      (NP-SBJ a jack-in-the-box))

(SINV (S-TPC-4 (NP-SBJ We)
               (VP will
                   (VP win)))
      ,
      (VP said
          (SBAR 0
                (S *T*-4))
          (NP-SBJ Casey)))
```

1.2.3 SBAR

SBAR is used for relative clauses and subordinate clauses, including indirect questions.

```
(S (NP-SBJ (NP The person)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP threw
                        (NP the ball)))))
   (VP is
       (ADJP-PRD very athletic)))

(S (NP-SBJ Willie)
   (VP knew
       (SBAR that
             (S (NP-SBJ Casey)
                (VP threw
```

```
                    (NP the ball))))))
```

```
(S (NP-SBJ Willie)
   (VP asked
       (SBAR (WHNP-1 who)
             (S (NP-SBJ *T*-1)
                (VP threw
                    (NP the ball))))))
```

The *wh*-prefixed labels, WHNP, WHADVP, WHADJP, WHPP, are used only when there is *wh*-movement, and they always leave a trace *T*. (See section 9 [*WH*-Phrases] for more information.)

```
(S (NP-SBJ-1 The committee)
   (VP continued
       (NP its meeting)
       (SBAR-TMP while
                 (S (NP-SBJ *-1)
                    (VP eating
                        (NP lunch))))))
```

1.2.4 RRC (reduced relative clause)

Reduced relative clauses are adjoined to the NP they modify.

```
(NP (NP An orangutan)
    (VP foaming
        (PP-CLR at
                (NP the mouth))))
```

All sentential modifiers in reduced relatives are attached at VP level, even when they precede the verb itself.

```
(NP (NP a car)
    (VP not built
        (NP *)
        (PP by
            (NP-LGS Mazda))
        (PP-TMP for
                (NP the last five years))))
```

The RRC label is used only in cases where there is no VP and an extra level is needed for proper attachment of sentential modifiers (see section 13 [Gerunds and Participles]).

```
(NP (NP 110 titles)
    (RRC not
         (ADVP-TMP presently)
         (PP-LOC in
                 (NP the collection))))
```

In the case of passives, a passive trace (NP *) is inserted under the VP, but it is not indexed, since the null subject to which the passive trace would be coindexed is not present in the annotation. The logical subject is tagged -LGS.

```
(S (NP-SBJ I)
   (VP bought
       (NP (NP a car)
           (VP built
               (NP *)
               (PP by
                   (NP-LGS Mazda))))))
```

Compare with the unreduced:

```
(S (NP-SBJ I)
   (VP bought
       (NP (NP a car)
           (SBAR (WHNP-5 which)
                 (S (NP-SBJ-6 *T*-5)
                    (VP was
                        (VP built
                            (NP *-6)
                            (PP by
                                (NP-LGS Mazda)))))))))
```

1.2.5 SBARQ

The SBARQ label marks *wh*-questions (i.e., those that contain a gap and therefore require a trace). A further level of structure, SQ, contains the inverted auxiliary (if there is one) and the rest of the sentence. The inverted auxiliary in *wh*-questions is not labeled.

```
(SBARQ (WHNP-1 Who)
       (SQ (NP-SBJ *T*-1)
           (VP threw
               (NP the ball)))
       ?)


(SBARQ (WHNP-2 What)
       (SQ did
           (NP-SBJ Casey)
           (VP throw
               (NP *T*-2)))
       ?)


(SBARQ (WHNP-3 Who)
       (SQ (NP-SBJ *T*-3)
           (VP will
               (VP throw
                   (NP the ball))))
       ?)
```

If the main verb is inverted, most arguments and adjuncts go at SQ level:

```
(SBARQ (WHADVP-1 Why)
       (SQ (PP-LOC in
```

```
                (NP these movies))
        is
        (NP-SBJ the unwed pregnant woman)
        (ADVP-TMP always)
        (PP-PRD from
                (NP Ohio))
        (ADVP-PRP *T*-1))
    ?)
```

1.2.6 SQ

(See also section 1.2.7.)

- inside SBARQ:

  As described above, inside *wh*-questions, SQ holds the subject, inverted auxiliary (if any), main verb phrase, and some adjuncts.

- yes/no questions:

  SQ is used for yes/no questions (i.e., those with inversion but no *wh*-movement).

```
(SQ Did
    (NP-SBJ Casey)
    (VP throw
        (NP the ball))
    ?)
```

- subject-less yes/no questions:

  In questions where the auxiliary and subject do not appear, the auxiliary is unlabeled and a null subject (NP-SBJ *) is used.

```
(SQ (NP-SBJ *)
    (VP See
        (NP that cute dog))
    ?)
```

  Note that questions with overt subjects and auxiliaries that show declarative word order are simply labeled S.

- Tag questions:

  Tag questions are treated as an adjunction of SQ to S. The resulting structure is labeled SQ, since the whole thing is interrogative in nature. The lower SQ is annotated to show predicate deletion; that is, an appropriate null *?* is inserted.

```
(SQ (S But
       (NP-SBJ you)
       (VP knew
           (NP that)))
    ,
    (SQ did n't
```

```
          (NP-SBJ you)
          (VP *?*))
      ?)

   (SQ (S (NP-SBJ That)
          (VP 's
              (NP-PRD the problem)))
       ,
       (SQ is
           n't
           (NP-SBJ it)
           (NP-PRD *?*))
       ?)
```

1.2.7 S-CLF

(*it*-cleft or "true" cleft)

Declarative *it*-clefts are labeled S-CLF, expletive *it* is tagged as the surface subject (-SBJ), the SBAR is attached at VP-level, and a trace is coindexed to the *wh*-complementizer of the clefted portion. (See section 16 [Clefts] for more information.)

```
(S-CLF (NP-SBJ It)
       (VP was
           (NP Casey)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP threw
                        (NP the ball))))))
```

Traces for all adverbials, including purpose clauses, are labeled (ADVP *T*) and coindexed to the WHADVP complementizer. If a preposition is pied-piped, WHPP and (PP *T*) are used.

```
(S-CLF (NP-SBJ It)
       (VP is
           (ADVP-TMP-PRD then)
           (SBAR (WHADVP-1 that)
                 (S (NP-SBJ-2 young queens)
                    (VP begin
                        (S (NP-SBJ *-2)
                           (VP to
                               (VP appear)))
                        (ADVP-TMP *T*-1))))))
```

Interrogative *it*-clefts are labeled SQ-CLF:

```
(SQ-CLF Was
        (NP-SBJ it)
        (NP-PRD John)
        (SBAR (WHNP-1 who)
              (S (NP-SBJ *T*-1)
                 (VP came
```

```
                    (PP-DIR to
                            (NP the party))
                    (PP-CLR in
                            (NP a dress)))))
         ?)
```

Note: *wh*-clefts do not receive special treatment in the corpus. They contain a free/headless relative, followed by a form of the verb *be*, followed by a predicate (e.g., *What sustained the worst injuries / was / the car*). (See section 2 [Notation] for a bit more on free relatives.)

1.2.8 *it*-extraposition

Clauses that are extraposed from subject position are labeled S or SBAR. The extraposed clause is attached at VP level and adjoined to the "it" with *EXP*-attach. The NP containing *it* and *EXP* is tagged -SBJ. (See section 17 [*It*-Extraposition] for more information.)

```
(S (NP-SBJ (NP It)
           (S *EXP*-1))
   (VP is
       (NP-PRD a pleasure)
       (S-1 (NP-SBJ *)
            (VP to
                (VP teach
                    (NP her))))))
```

1.2.9 FRAG

(fragment)

FRAG marks those portions of text that appear to be clauses, but lack too many essential elements for the exact structure to be easily determined (e.g., answers to questions). Predicate argument structure therefore cannot be extracted from FRAGs. Some examples of what we have called FRAG:

```
(SBARQ (WHNP-9 Who)
       (S (NP-SBJ *T*-9)
          (VP threw
              (NP the ball)))
       ?)
(FRAG (NP Casey)
      ,
      (NP-TMP yesterday))


(S (SBAR-ADV if
             (FRAG not
                   (NP-TMP today)))
   ,
   Casey will throw the ball tomorrow)


(SBAR-ADV Though
          (FRAG (ADJP limited)))
```

```
(SBAR-ADV if
          (FRAG (ADJP possible)))

(SBAR-ADV (WHNP Whatever)
          (FRAG (NP the long-term economic effect)))

(FRAG (PP-LOC Among
              (NP (NP the Guinness disk 's)
                  wonders))
      :
      (NP (NP the world 's)
          loudest recorded belch)
      .)

( (FRAG (NP (NP Two guys)
            (PP from
                (NP (NP Gary)
                    ,
                    (NP Ind.))))
        ?
        ''))

( (FRAG Not
        (ADJP so)
        .))

( (FRAG (VP Guaranteed
            (NP *)
            (PP by
                (NP-LGS India)))
        .))
```

1.3 Clause combinations

1.3.1 Coordination

See section 8 [Shared Complements and Modifiers] and section 7 [Coordination] for details concerning coordination structures.

- Phrase coordination

  Coordination of phrases is represented in the annotation at the lowest level possible. Single words are assumed to coordinate at word level rather than projecting their own phrases, and only the highest level is represented.

```
(S (NP-SBJ Girls and boys)
   (VP throw and catch
       (NP balls)))
```

However, the addition of modifiers generally forces a higher level of coordination, which is shown with Chomsky-adjunction structure.

```
(S (NP-SBJ (NP These girls)
           and
           (NP those boys))
   (VP (VP throw
           (ADVP-MNR well))
       and
       (VP catch
           (ADVP-MNR badly))))
```

- Clause coordination

When like clauses are coordinated, the level of coordination has the same label as the coordinated clauses.

```
(S (S (NP-SBJ Casey)
      (VP threw
          (NP the ball)))
   and
   (S (NP-SBJ Willie)
      (VP caught
          (NP it))))
```

```
(S (NP-SBJ Jackie)
   (VP knew
       (SBAR (SBAR (WHNP-1 who)
                   (S (NP-SBJ *T*-1)
                      (VP threw
                          (NP the ball))))
             and
             (SBAR (WHNP-2 who)
                   (S (NP-SBJ *T*-2)
                      (VP caught
                          (NP it)))))))
```

Coordinated unlike clauses are dominated by S.

```
(S (S (NP-SBJ He)
      (VP was
          (VP performing
              (PP for
                  (NP pay)))))
   , and
   (SBARQ (WHADVP-1 why)
          (SQ should
              (NP-SBJ anyone)
              (VP expect
                  (NP (NP anything)
                      (ADJP more))
```

```
                  (ADVP-PRP *T*-1))))
   ?)
```

Note that coordination of unlike phrases (i.e., non-clauses) is dominated by UCP (Unlike Coordinated Phrase).

- Coordinating conjunctions

  Besides the usual *and, or, but*, etc., certain prepositions and subordinating conjunctions can be used as coordinating conjunctions. Multi-word coordinating conjunctions are labeled CONJP (see section 7 [Coordination]).

```
(S (NP-SBJ Willie
           (CONJP as well as)
           Casey)
   (VP saw
       (NP the ball)))


(S (NP-SBJ Casey)
   (VP will
       (VP throw
           (NP the ball)
           (NP-TMP tomorrow
                   (CONJP if not)
                   today))))
```

  Otherwise, these are annotated as PPs or SBARs, as appropriate. (See section 7 [Coordination].)

```
(S (SBAR-ADV if
             (FRAG not
                   (NP today)))
   ,
   (NP-SBJ Casey)
   (VP will
       (VP throw
           (NP the ball)
           (NP-TMP tomorrow))))
```

1.3.2 Subordination: the use of SBAR

See section 9 [*WH*-Phrases] and section 10 [Subordinate Clauses] for more detail.

- Relative clauses are adjoined to the NP that they modify.

```
(S (NP-SBJ (NP The person)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP caught
                        (NP the ball)))))
   (VP is
       (ADJP-PRD very athletic)))
```

- SBARs can be verbal complements.

```
(S (NP-SBJ Willie)
   (VP knew
       (SBAR that
             (S (NP-SBJ Casey)
                (VP hid
                    (NP the ball))))))
```

- Conditional, temporal, and other such adverbial SBARs are attached under either S or VP, depending on whether they precede or follow the main clause, and given the appropriate adverbial function tag.

```
(S (SBAR-ADV if
             (S (NP-SBJ Casey)
                (VP throws
                    (NP the ball))))
   (NP-SBJ Willie)
   (VP catches
       (NP it)))


(S (NP-SBJ Willie)
   (VP catches
       (NP the ball)
       (SBAR-ADV if
                 (S (NP-SBJ Casey)
                    (VP throws
                        (NP it))))))
```

1.3.3 Fronted elements

Fronted elements are those that appear before the subject in a declarative sentence. They are placed inside the top clause level (e.g. S, SINV, SQ, SBAR). (See the section on -TPC in section 2 [Notation] for more details on the use of the -TPC tag and the section on *T* with fronted elements in section 4 [Null Elements] for more details on the distribution of *T*.)

- Fronted arguments.

  Fronted arguments are attached under the main clause level. They always leave a *T* and are tagged -TPC. This holds whether the argument is fronted within a single clause or crosses more than one clause boundary.

```
(S (NP-TPC-5 This)
   (NP-SBJ every man)
   (VP contains
       (NP *T*-5)
       (PP-LOC within
               (NP him))))
```

- Fronted adjuncts.

  Fronted adjuncts receive function tags (e.g, -ADV, -TMP) as appropriate. However, note that they receive the -TPC label only in cases where they are associated with a *T* in a lower clause. In particular, a *T* appears in the annotation only if the adjunct is fronted over more than one clause boundary:

```
(S (PP-TPC-1 Excluding
             (NP (NP (NP an increase)
                     (PP-LOC in
                             (NP the tax rate)))
                 and
                 (NP (NP the effects)
                     (PP of
                         (NP foreign currency translations)))))
   ,
   (NP-SBJ Mr. Millis)
   (VP said
       (SBAR 0
             (S (NP-SBJ (NP the company 's)
                        results)
                ``
                (VP were
                    (ADVP still)
                    (ADJP-PRD (NP-ADV a little)
                              disappointing)
                    (PP *T*-1)))))
   . '')
```

```
(S (SBAR-ADV-TPC-1 If
                   (S (NP-SBJ profits)
                      (VP do n't
                          (VP improve))))
   ,
   (NP-SBJ Mr. Whitten)
   (VP says
       (SBAR 0
             (S (NP-SBJ he)
                (VP may
                    (VP quit
                        (NP the exchange)
                        (SBAR-ADV *T*-1))))))
   .)
```

  There is no *T* in the annotation if the adjunct is fronted within a single clause:

```
(S (NP-TMP Yesterday)
   (NP-SBJ I)
   (VP went
       (PP-DIR to
               (NP the store))))
```

```
(S (S-ADV (NP-SBJ *-1)
          (VP Running
              (PP-DIR toward
                      (NP Casey))))
   ,
   (NP-SBJ-1 Willie)
   (VP caught
       (NP the ball)))


(S (SBAR-TMP (WHADVP-7 When)
             (S (NP-SBJ I)
                (VP do n't
                    (VP get
                        (NP enough sleep)
                        (ADVP-TMP *T*-7)))))
   (NP-SBJ-8 I)
   (VP have
       (NP (NP trouble)
           (S-NOM (NP-SBJ *-8)
                  (VP staying
                      (ADJP-PRD awake))))))
```

Elements that are fronted within questions (either *wh-* or yes/no) are put inside the highest level of structure (SBARQ or SQ, respectively).

```
(SBARQ (SBAR-ADV If
                 (S (NP-SBJ Casey)
                    (VP throws
                        (NP the ball))))
       (WHNP-1 who)
       (SQ will
           (NP-SBJ *T*-1)
           (VP catch
               (NP it)))
       ?)


(SQ (SBAR-ADV If
              (S (NP-SBJ Casey)
                 (VP throws
                     (NP the ball))))
    will
    (NP-SBJ Willie)
    (VP catch
        (NP it))
    ?)
```

1.3.4 Quotations

- Direct quotations

  A direct quotation is considered to be the argument of the verb of saying.

```
(S (NP-SBJ Casey)
   (VP said
       ``
       (S (NP-SBJ Willie)
          (VP caught
              (NP the ball))))
   .
   '')
```

When the quotation appears before the quoting verb, it treated as a fronted argument: the quote is attached at S level and given a -TPC tag, and a trace is shown under the VP.

```
(S ``
   (S-TPC-1 (NP-SBJ Willie)
            (VP caught
                (NP the ball)))
   ,
   ''
   (NP-SBJ Casey)
   (VP said
       (S *T*-1))
   .)
```

```
(SINV ``
      (S-TPC-1 (NP-SBJ Willie)
               (VP caught
                   (NP the ball)))
      ,
      ''
      (VP said
          (S *T*-1))
      (NP-SBJ Casey)
      .)
```

If the quotation is discontinuous, the interruptive material is annotated as a parenthetical (-PRN). Note that a trace appears under the VP in the parenthetical, but that the fronted portion is not labeled -TPC:

```
(S-7 ``
     (NP-SBJ Willie)
     ''
     (PRN ,
          (S (NP-SBJ Casey)
             (VP said
                 (S *T*-7)))
          ,)
     ``
     (VP caught
         (NP the ball))
     .
     '')
```

- Indirect quotations

  The (SBAR 0) level marks indirect quotations and cases where only a portion of the quote is direct.

```
(S (NP-SBJ Stokely)
   (VP says
       (SBAR 0
             (S (NP-SBJ stores)
                (VP revive
                    (NP (NP specials)
                        (PP like
                            (NP (NP three cans)
                                (PP of
                                    (NP peas))
                                (PP for
                                    (NP 99 cents))))))))
   .)
```

```
(S (NP-SBJ Mr. Millis)
   (VP said
       (SBAR 0
             (S (NP-SBJ (NP the company 's)
                        results)
                ``
                (VP were
                    (ADVP still)
                    (ADJP-PRD (NP-ADV a little)
                              disappointing)))))
   . '')
```

1.3.5 Gapping

See section 7 [Coordination] for more details on gapping.

- Intrasentential gapping

  When the gapped construction exists alongside a complete (ungapped) clause of parallel structure, the complete clause is used as a template to which elements in the gapped clause are mapped via the "=" notation (referred to as "gap coindexing").

```
(S (S (NP-SBJ-1 Mary)
      (VP likes
          (NP-2 Bach)))
   and
   (S (NP-SBJ=1 Susan)
      ,
      (NP=2 Beethoven)))
```

  In the above, the equal sign notation maps constituent NP=1 over NP-1 so that the following predicate argument structure can be extracted: LIKE (Mary, Bach) and LIKE (Susan, Beethoven). Similarly:

```
(S (NP-SBJ John)
   (VP (VP gave                  => GIVE (John, Mary, book)
           (NP-1 Mary)
           (NP-2 a book))
       and
       (VP (NP=1 Bill)           => GIVE (John, Bill, pencil)
           (NP=2 a pencil))))
```

All constituents in gapped constructions receive function tags as appropriate.

```
(S (NP-SBJ I)
   (VP (VP eat
           (NP-1 breakfast)
           (PP-TMP-2 in
                     (NP the morning)))
       and
       (VP (NP=1 lunch)
           (PP-TMP=2 in
                     (NP the afternoon)))))
```

- Intersentential gapping

  The template approach to gapping is not used across sentences. If a template is not available within the same sentence, the gapped constituent is labeled FRAG instead.

  What is Tim eating? Mary Ann thinks chocolate.

```
(S (NP-SBJ Mary Ann)
   (VP thinks
       (SBAR 0
             (FRAG (NP chocolate)))))
```

2 Notation

In the present corpus, each bracket is labeled for at least 1 syntactic category but may have as many as 4 function tags. In previous Treebank releases, only standard syntactic labels (e.g. NP, ADVP, PP, etc.) were used to label constituents; every bracket had just one label. The limitations of this system become apparent when a word or phrase that belongs to one syntactic category is used for some other function or when it plays a role that is not easily identified without special annotation.

In addition to the function tags, we have also augmented our annotation with the coindexation of null elements and several varieties of "pseudo-attach".

2.1 Bracket labels

2.1.1 Clause level

See section 1 [Overview of Basic Clause Structure] for a fuller explanation of the use(s) of each of these labels.

S — Simple declarative clause, i.e. one that is not introduced by a (possibly empty) subordinating conjunction or *wh*-word and that does not exhibit subject-verb inversion.

SBAR — Clause introduced by a (possibly empty) subordinating conjunction.

SBARQ — Direct question introduced by a *wh*-word or *wh*-phrase. See section 1 [Overview of Basic Clause Structure]. Indirect questions and relative clauses should be bracketed as SBAR, not SBARQ.

SINV — Inverted declarative sentence, i.e. one in which the subject follows the tensed verb or modal.

SQ — Inverted yes/no question, or main clause of a *wh*-question, following the *wh*-phrase in SBARQ.

2.1.2 Phrase level

ADJP — Adjective Phrase. Phrasal category headed by an adjective (including comparative and superlative adjectives). Example: *outrageously expensive*.

ADVP — Adverb Phrase. Phrasal category headed by an adverb (including comparative and superlative adverbs). Examples: *rather timidly*, *very well indeed*, *rapidly*.

CONJP — Conjunction Phrase. Used to mark certain "multi-word" conjunctions, such as *as well as, instead of.*

FRAG — Fragment. (see section 1 [Overview of Basic Clause Structure])

INTJ — Interjection. Corresponds approximately to the part-of-speech tag UH (see the *POS guidelines* [Santorini 1990]).

LST — List marker. Includes surrounding punctuation. (see section 24 [Numbered Lists])

NAC — Not A Constituent; used to show the scope of certain prenominal modifiers within a noun phrase.

NP — Noun Phrase. Phrasal category that includes all constituents that depend on a head noun.

NX — Used within certain complex noun phrases to mark the head of the noun phrase. Corresponds very roughly to N-bar level but used quite differently.

PP — Prepositional Phrase. Phrasal category headed by a preposition.

PRN — Parenthetical. (See section 2.6 below for a complete description.)

PRT — Particle. Category for words that should be tagged RP, as described in the *POS guidelines* [Santorini 1990], with some guidance from [Quirk et al. 1985] sections 16.3-16 in tricky ADVP vs. PRT decisions (but note that the Treebank notion of particle is somewhat different from that of Quirk et al.).

QP — Quantifier Phrase (i.e., complex measure/amount phrase); used within NP. (see section 11 [Modification of NP])

RRC — Reduced Relative Clause. (see section 13 [Gerunds and Participles])

UCP — Unlike Coordinated Phrase.

VP — Verb Phrase. Phrasal category headed a verb.

WHADJP — *Wh*-adjective Phrase. Adjectival phrase containing a *wh*-adverb, as in *how hot*.

WHADVP — *Wh*-adverb Phrase. Introduces a clause with an ADVP gap. May be null (containing the `0` complementizer) or lexical, containing a *wh*-adverb such as *how* or *why*.

WHNP — *Wh*-noun Phrase. Introduces a clause with an NP gap. May be null (containing the `0` complementizer) or lexical, containing some *wh*-word, e.g. *who*, *which book*, *whose daughter*, *none of which*, or *how many leopards*.

WHPP — *Wh*-prepositional Phrase. Prepositional phrase containing a *wh*-noun phrase (such as *of which* or *by whose authority*) that either introduces a PP gap or is contained by a WHNP.

X — Unknown, uncertain, or unbracketable. X is often used for bracketing typos and in bracketing *the...the*-constructions (see section 10 [Subordinate Clauses] and section 25 [Correlative *the*-Clauses]).

2.2 Function tags

2.2.1 Form/function discrepancies

-ADV (adverbial) — marks a constituent other than ADVP or PP when it is used adverbially (e.g., NPs or free ("headless") relatives). However, constituents that themselves are modifying an ADVP generally do not get -ADV.

```
(ADJP (NP-ADV a little bit)
      angry)


(S (NP-SBJ You)
   (VP can
       (VP leave
           (SBAR-ADV if
                     (S (NP-SBJ-8 you)
                        (ADVP really)
                        (VP want
                            (S (NP-SBJ *-8)
                               (VP to
                                   (VP go)))))))))
```

If there is a more specific adverbial tag available (i.e. one of the tags listed in section 2.2.3 below), the more specific tag is assumed to imply -ADV and is used alone. Thus in the following example, the -TMP tag on *yesterday* implies -ADV.

```
(S (NP-SBJ He)
   (VP left
       (NP-TMP yesterday)))
```

NOT:

```
(NP-ADV-TMP yesterday)
```

Nouns such as *today*, which often behave adverbially, are labeled NP when they appear in argument position.

```
(S (NP-SBJ Today)
   (VP is
       (NP-PRD (NP the first day)
               (PP of
                   (NP (NP the rest)
                       (PP of
                           (NP your life)))))))
```

-NOM (nominal) — marks free ("headless") relatives and gerunds when they act nominally. (See section 9 [*WH*-Phrases] for more information about free relatives, and section 13 [Gerunds and Participles] for more information about gerunds.)

```
(S (SBAR-NOM-SBJ (WHNP-10 What)
                 (S (NP-SBJ I)
                    (ADVP really)
                    (VP like
                        (NP *T*-10))))
   (VP is
       (NP-PRD chocolate)))


(S (S-NOM-SBJ (NP-SBJ *)
              (VP Baking
                  (NP pies)))
   (VP is
       (ADJP-PRD fun)))


(S (NP-SBJ I)
   (VP do
       not
       (VP mind
           (PP about
               (S-NOM (NP-SBJ your)
                      (VP leaving
                          (ADVP-TMP early)))))))
```

Note that other non-NP constituents are not tagged -NOM when they appear in argument positions (e.g., infinitivals, PPs), though they do get -SBJ when they occur in subject position.

```
(S (S-SBJ (NP-SBJ *)
          (VP To
              (VP have
                  (VP refused))))
   (VP would
       (VP have
           (VP been
               (NP-PRD political suicide)))))
```

```
(S (PP-SBJ (PP From
               (NP proud pool-owners))
           (PP to
               (NP perpetual hosts and handymen)))
   (VP was
       (NP-PRD a short step)))
```

2.2.2 Grammatical role

-DTV (dative) — marks the dative object in the unshifted form of the double object construction.

If the preposition introducing the “dative” object is *for*, it is considered benefactive (see -BNF on page 41).

```
(S (NP-SBJ I)
   (VP asked
       (NP a question)
       (PP-DTV of
               (NP the president))))
```

```
(S (NP-SBJ Aristotle)
   (VP gave
       (NP the book)
       (PP-DTV to
               (NP Plato))))
```

Compare with the shifted

```
(S (NP-SBJ Aristotle)
   (VP gave
       (NP Plato)
       (NP the book)))
```

Other verbs have semantically similar complements that could be considered “dative” objects. However, -DTV (or -BNF) is only used after verbs that can undergo dative shift. Other putative datives are annotated with -CLR.

```
(S (NP-SBJ He)
   (VP donated
       (NP money)
```

```
        (PP-CLR to
                (NP the museum))))
```

-LGS (logical subject) — is used to mark the logical subject in passives. It attaches to the NP object of *by* and not to the PP node itself.

```
(S (NP-SBJ-7 That)
   (VP was
       (VP painted
           (NP *-7)
           (PP by
               (NP-LGS Mark)))))
```

-PRD (predicate) — marks any predicate that is not VP.

```
(S (NP-SBJ I)
   (VP consider
       (S (NP-SBJ Kris)
          (NP-PRD a fool))))
```

```
(SQ Was
    (NP-SBJ he)
    (ADVP-TMP ever)
    (ADJP-PRD successful)
    ?)
```

In *do so* constructions, the *so* is annotated as a predicate.

```
(S (NP-SBJ They)
   (ADVP also)
   (VP did
       (ADVP-PRD so)))
```

```
(SINV and
      (ADVP-PRD-TPC-1 so)
      (VP did
          (ADVP-PRD *T*-1))
      (NP-SBJ the hippopotamuses))
```

-PUT — marks the locative complement of *put*.

```
(S (NP-SBJ John)
   (VP put
       (NP the book)
       (PP-PUT on
               (NP the table))))
```

```
(S (NP-SBJ John)
   (VP put
       (NP the book)
       (ADVP-PUT there)))
```

It does **not** go on just any complement or child of *put*:

```
(S (NP-SBJ They)
   (VP put
       (NP the baby)
       (PRT down)))
```

```
(S (NP-SBJ She)
   (VP put
       (NP it)
       (ADVP-MNR bluntly)))
```

-SBJ (surface subject) — marks the structural surface subject of both matrix and embedded clauses, including those with null subjects.

-TPC ("topicalized") — marks elements that appear before the subject in a declarative sentence, but in two cases only:

(i) if the fronted element is associated with a *T* in the position of the gap.

(ii) if the fronted element is left-dislocated (i.e., it is associated with a resumptive pronoun in the position of the gap). (See the section on fronted elements in section 1 [Overview of Basic Clause Structure] for more details on the treatment of fronted elements and the section on *T* with fronted elements in section 4 [Null Elements] for more details on the distribution of *T*.)

```
(S (PP-TPC-12 Of
              (NP (NP the 500 barbers)
                  (PP-LOC in
                          (NP Philadelphia))))
   ,
   (NP-SBJ (NP (QP only 10))
           (PP *T*-12))
   (VP know
       (SBAR (WHNP-13 what)
             (S (NP-SBJ they)
                (VP are
                    (VP doing
                        (NP *T*-13)))))))
```

-VOC (vocative) — marks nouns of address, regardless of their position in the sentence. It is not coindexed to the subject and does not get -TPC when it is sentence-initial.

```
(SQ (NP-VOC Mike)
    ,
    would
    (NP-SBJ you)
```

```
(INTJ please)
(VP close
    (NP the door))
?)
```

2.2.3 Adverbials

Adverbials are generally VP adjuncts.

-BNF (benefactive) — marks the beneficiary of an action (attaches to NP or PP).

This tag is used *only* when (1) the verb can undergo dative shift and (2) the prepositional variant (with the same meaning) uses *for*. The prepositional objects of dative-shifting verbs with other prepositions than *for* (such as *to* or *of*) are annotated -DTV.

```
(S (NP-SBJ I)
   (VP baked
       (NP-BNF Doug)
       (NP a cake)))
```

```
(S (NP-SBJ I)
   (VP baked
       (NP a cake)
       (PP-BNF for
               (NP Doug))))
```

-DIR (direction) — marks adverbials that answer the questions "from where?" and "to where?" It implies motion, which can be metaphorical as in *"...rose 5 pts. to 57-1/2"* or *"increased 70% to 5.8 billion yen"* (see section 23 ["Financialspeak" Conventions]). -DIR is most often used with verbs of motion/transit and financial verbs:

```
(S (NP-SBJ I)
   (VP flew
       (PP-DIR from
               (NP Tokyo))
       (PP-DIR to
               (NP New York))))
```

-EXT (extent) — marks adverbial phrases that describe the spatial extent of an activity. -EXT was incorporated primarily for cases of movement in financial space, but is also used in analogous situations elsewhere.

```
(S (NP-SBJ the Dow Jones Industrial Average)
   (VP plunged
       (NP-EXT 190.58 points)))
```

```
(S (NP-SBJ She)
   (VP walked
       (NP-EXT 5 miles)))
```

Obligatory complements do not receive -EXT:

```
(S (NP-SBJ The sumo wrestler)
   (VP gained
       (NP 80 pounds)))
```

Words such as *fully* and *completely* are absolutes and do **not** receive -EXT.

-LOC (locative) — marks adverbials that indicate place/setting of the event.

```
(PP-LOC on
        (NP the moon))
```

-LOC may also indicate metaphorical location. For example, the following receive the -LOC tag:

```
(PP-LOC amongst
        (NP yourselves))
```

```
(NP (NP a drop)
    (PP-LOC in domestic truck sales))
```

whereas these do not:

```
(NP (NP interest)
    (PP in anthropology))
```

```
(NP (NP 5 dollars)
    (PP in
        (NP stocks)))
```

```
(PP on
    (NP (NP grounds)
        (PP of ...)))
```

```
(PP in
    (NP other respects))
```

```
(PP under
    (NP pressure))
```

```
(PP under
    (NP the gun))
```

There is likely to be some variation in the use of -LOC due to differing annotator interpretations. In cases where the annotator is faced with choosing between -LOC or -TMP, the DEFAULT is -LOC:

```
(PP-LOC in
        (NP an interview))
```

```
(PP-LOC in
        (NP active trading))
```

In cases of apposition involving SBAR, the SBAR should not be labeled -LOC.

```
(NP (NP Minneapolis)
    ,
    (SBAR (WHADVP-1 where)
          (S (NP-SBJ it)
             (VP is
                 (ADJP-PRD cold)
                 (ADVP-LOC *T*-1)))))
```

-LOC has some uses that are not adverbial, such as with place names that are adjoined to other NPs and NAC-LOC premodifiers of NPs (see section 11 [Modification of NP]). The special tag -PUT, listed on page 39, is used for the locative argument of *put*.

-MNR (manner) — marks adverbials that indicate manner, including instrument phrases.

```
(S (NP She)
   (VP waited
       (ADVP-MNR impatiently)))
```

```
(S (NP-SBJ She)
   (VP hit
       (NP the nail)
       (PP-MNR with
               (NP a hammer))))
```

```
(S (NP-SBJ-14 She)
   (VP surprised
       (NP him)
       (PP-MNR by
               (S-NOM (NP-SBJ *-14)
                      (VP eating
                          (NP a horse)
                          (ADVP-MNR alone))))))
```

-PRP (purpose or reason) — marks purpose or reason clauses and PPs.

```
(S (NP-SBJ-1 Chevron)
   (VP had
       (S (NP-SBJ *-1)
          (VP to
              (VP shut
```

```
                  (PRT down)
                  (NP (NP a crude-oil pipeline)
                      (PP-LOC in
                              (NP the Bay area))))))
       (S-PRP (NP-SBJ *-1)
              (VP to
                  (VP check
                      (PP-CLR for
                              (NP leaks)))))))
```

```
(S (NP-SBJ the Dow Jones Transportation Average)
   (VP went
       (ADVP-DIR down)
       ,
       (PP-PRP due
               (ADVP largely)
               (PP to
                   (NP (NP further selling)
                       (PP-LOC in
                               (NP UAL)))))))
```

```
(S (NP-SBJ-1 (NP activity)
             (PP-LOC at
                     (NP (NP a number)
                         (PP of
                             (NP (ADJP San Francisco-based)
                                 brokerage houses)))))
   (VP was
       (VP curtailed
           (NP *-1)
           (PP-PRP as
                   (NP (NP a result)
                       (PP of
                           (NP the earthquake)))))))
```

-TMP (temporal) — marks temporal or aspectual adverbials that answer the questions *when*, *how often*, or *how long*. It has some uses that are not strictly adverbial, such as with dates that modify other NPs (see section 11 [Modification of NP]).

at S- or VP-level:

```
(S (NP-SBJ Egg bread)
   (VP loses
       (NP some zip)
       (SBAR-TMP (WHADVP-2 when)
                 (S (NP-SBJ the eggs)
                    (VP come
                        (PP-LOC-CLR in
                                    (NP 30-pound cans)
                                    (ADVP-TMP *T*-2)))))))
```

```
(S (ADVP-TMP Meanwhile)
   ,
   (NP-SBJ (NP the bottom end)
           (PP of
               (NP the market)))
   (VP is
       (VP becoming
           (ADJP-PRD less loyal))))
```

```
(S (NP-SBJ Brand loyalty)
   (VP has
       (VP eroded
           (PP-TMP during
                   (NP the 1980s)))))
```

```
(S (NP-SBJ-2 it)
   (VP will
       (VP remove
           (NP the objectionable tropical oil)
           (PP-TMP by
                   (NP year end)))))
```

modifying NPs:

```
(S (NP-SBJ the 26-man Politburo)
   (VP had
       (VP asked
           (PP-CLR for
                   (NP his resignation))
           (PP-LOC at
                   (NP (NP a separate meeting)
                       (NP-TMP late Tuesday))))))
```

```
(NP (NP his (ADJP first and only) state visit)
    (PP to
        (NP Bonn))
    (ADVP-TMP (NP two years)
              ago))
```

In cases of apposition involving SBAR, the SBAR should not be labeled -TMP.

```
(PP-TMP in
        (NP (NP 1992)
            ,
            (SBAR (WHADVP-4 when)
                  (S (NP-SBJ I)
                     (ADVP-TMP first)
                     (VP learned
                         (SBAR 0
```

```
(S (NP-SBJ I)
   (VP was
       (NP-PRD a Martian)
       (ADVP-TMP *T*-4))))))))
```

in ADJP:

```
(NP (NP (NP Penn Treebank Corp. 's)
        (ADJP 14.7 %)
        bonds)
    (ADJP due
          (NP-TMP 2009)))
```

Only in "financialspeak," and only when the dominating PP is a PP-DIR, may temporal modifiers be put at PP object level, as in this example:

```
(S (NP-SBJ unconsolidated pretax profit)
   (VP increased
       (PP-DIR to
               (NP (QP 12.12 billion) yen))
       ,
       (PP-DIR from
               (NP (QP 7.12 billion) yen)
               (ADVP-TMP (NP a year)
                         ago))))
```

Note that -TMP is not used in possessive phrases:

```
(NP (NP 1950 's)
    conservative tendencies)
```

2.2.4 Miscellaneous

-CLR (closely related) — marks constituents that occupy some middle ground between argument and adjunct of the verb phrase. These roughly correspond to "predication adjuncts" , prepositional ditransitives , and some "phrasal verbs" , as defined in [Quirk et al. 1985].

Although constituents marked with -CLR are not strictly speaking complements, they are treated as complements whenever it makes a bracketing difference (see the section on fronting in section 1 [Overview of Basic Clause Structure] and *RNR*-attach in section 8 [Shared Complements and Modifiers]).

The precise meaning of -CLR depends somewhat on the category of its phrase:

on S or SBAR — These categories are usually arguments, so the -CLR tag indicates that the clause is more adverbial than normal clausal arguments. The most common case is the infinitival semi-complement of *use*, but there are a variety of other cases (see section 14 [Infinitives]).

on PP, ADVP, SBAR-PRP, etc. — On categories that are ordinarily interpreted as (adjunct) adverbials, -CLR indicates a somewhat closer relationship to the verb. For example:

- Prepositional Ditransitives
  In order to ensure consistency, the Treebank recognizes only a limited class of verbs that take more than one complement, as described in this section (-DTV on page 38 and -PUT on page 39) and in section 15 [Small Clauses]. Verbs that fall outside these classes (including most of the prepositional ditransitive verbs in class [D2] in [Quirk et al. 1985]) are often annotated with -CLR:

```
(VP associate
    (NP snow)
    (PP-CLR with
            (NP winter)))

(VP donate
    (NP your time)
    (PP-CLR to
            (NP a good cause)))
```

- Phrasal verbs
  Phrasal verbs are also annotated with -CLR or a combination of PRT and PP-CLR.

```
(VP pay
    (PP-CLR for
            (NP 500 shares)))

(VP adjust
    (PP-CLR for
            (NP inflation)))

(VP put
    (PRT up)
    (PP-CLR with
            (NP the nuisance)))
```

  Words that are considered borderline between particle and adverb are often bracketed with ADVP-CLR.

```
(VP looking
    (ADVP-CLR forward)
    (PP-CLR to
            (NP their winter meeting)))
```

- Predication Adjuncts
  Many of Quirk's predication adjuncts (see [Quirk et al. 1985], especially sections 8.27–35, 15.22, & 16.48) are annotated with -CLR.

```
(VP place
    (NP the flour)
    (PP-LOC-CLR in
                (NP the sifter)))

(S (NP-SBJ She)
   (VP sat
       (PP-LOC-CLR in
                   (NP the chair))))
```

```
(S (NP-SBJ He)
   (VP kissed
       (NP his mother)
       (PP-CLR on
               (NP the cheek))))
```

on NP — To the extent that -CLR is used on noun phrases, it indicates that the NP is part of some kind of "fixed phrase" or expression, such as *take care of*.

```
(VP taking
    (NP-CLR care)
    (PP-CLR of
            (NP the problem)))


(S (NP-SBJ their meeting)
   (ADVP-TMP never)
   (VP took
       (NP-CLR place)))
```

Variation is more likely for NPs than for other uses of -CLR.

-CLF (cleft) — marks it-clefts ("true" clefts) and may be added to the labels S, SINV, or SQ. See section 16 [Clefts].

```
(SQ-CLF Was
        (NP-SBJ it)
        (NP-PRD (NP John's)
                car)
        (SBAR (WHNP-6 0)
              (S (NP-SBJ you)
                 (VP borrowed
                     (NP *T*-6))))
        ?)
```

-HLN (headline) — marks headlines and datelines. Note that headlines and datelines always constitute a unit of text that is structurally independent from the following sentence.

```
( (NP-HLN (NP The end)
          (PP of
              (NP Trujillo))))
( (S (NP-SBJ (NP Assassination)
             (PRN ,
                  (PP (ADVP even)
                      of
                      (NP a tyrant))
                  ,))
     (VP is
         (ADJP-PRD repulsive
                   (PP to
                       (NP (NP men)
                           (PP of
                               (NP good conscience))))))
```

```
        .))

    ( (NP-HLN (NP-LOC Chicago , IL)
              ,
              (NP-TMP May 8)
              --))
    ( (S A fire broke out in an abandoned building .))
```

-TTL (title) — is attached to the top node of a title when this title appears inside running text. -TTL implies -NOM. The internal structure of the title is bracketed as usual. (See section 12 [Titles] for more information about the bracketing of titles.)

```
    (SQ Have
        (NP-SBJ you)
        (VP read
            (S-TTL (NP-SBJ *)
                   (VP To
                       (VP Kill
                           (NP A Mockingbird)))))
        ?)


    (S ``
       (NP-TTL-SBJ-1 Omphalos)
       ''
       (VP was
           (VP painted
               (NP *-1)
               (PP by
                   (NP-LGS Mark)))))
```

2.3 Null elements

See section 4 [Null Elements] for more on the annotation of null elements.

T — trace of A′-movement

(NP *) — arbitrary PRO, controlled PRO, and trace of A-movement

0 — the null complementizer

U — unit

? — placeholder for ellipsed material

NOT — anti-placeholder in template gapping

The "pseudo-attach" elements (listed in section 2.5) are also essentially null elements.

2.4 Coindexing

See section 4 [Null Elements] for more detailed information.

2.4.1 The identity index

The number that follows a *bracket tag* serves as an *identity index* (ID number) for that constituent: (WHNP-1 What). Identity indices only appear when necessary, e.g., when there is a corresponding null element.

2.4.2 The reference index

The number that follows the **null element** is called its *reference index*: (NP *T*-1). It should correspond to the identity index of the constituent with which the null is associated. If the null is not associated with a constituent in the same sentence, it does not receive an index.

```
(S (NP-SBJ Willie)
   (VP knew
       (SBAR (WHNP-1 who)
             (S (NP-SBJ *T*-1)
                (VP threw
                    (NP the ball))))))


(SBARQ (WHNP-1 Who)
       (SQ was
           (NP-SBJ-2 *T*-1)
           (VP believed
               (S (NP-SBJ-3 *-2)
                  (VP to
                      (VP have
                          (VP been
                              (VP shot
                                  (NP *-3))))))))
       ?)
```

2.5 Pseudo-attach

Pseudo-attach is a method of showing that non-adjacent constituents are related. There are four different types of pseudo-attach, each of which is described in detail in section 5 [Pseudo-Attach]. The pseudo attach copy bears a reference index corresponding to the identity index of the displaced constituent. The different types of pseudo-attach are:

EXP — Expletive (extraposition)

ICH — Interpret Constituent Here (discontinuous dependency)

PPA — Permanent Predictable Ambiguity (ambiguity)

RNR — Right Node Raising (shared complements)

2.6 Parentheticals

Parenthetical elements are dominated by a node labeled PRN. Punctuation marks that set off a parenthetical (i.e., commas, dashes, parentheses (-LRB- and -RRB-)) are contained within the PRN node. Use of PRN is determined ultimately by individual annotator intuition, though the presence of dashes or parentheses strongly suggests a parenthetical.

```
(S (NP-SBJ (NP Assassination)
           (PRN ,
                (PP (ADVP even)
                    of
                    (NP a tyrant))
                ,))
   (VP is
       (ADJP-PRD repulsive)
       (PP to
           (NP (NP men)
               (PP of
                   (NP good conscience))))))
```

```
(S (NP-TMP (NP Every day)
           (SBAR (WHNP-1 0)
                 (S (NP-SBJ you)
                    (VP delay
                        (NP-TMP *T*-1)))))
   ,
   (NP-SBJ (NP (NP a savings institution 's) health)
           (PRN --
                and
                (NP the federal budget deficit)
                --))
   (VP grows
       (ADJP-PRD worse)))
```

```
(S (NP-SBJ Casey)
   (VP threw
       (NP the
           (ADJP red
                 (PRN (S (ADVP at least)
                         (NP-SBJ I)
                         (VP think
                             (SBAR 0
                                   (S (NP-SBJ it)
                                      (VP was
                                          (ADJP-PRD red)))))))
                 and
                 green
                 (PRN or
                      (S (ADVP maybe)
                         (NP-SBJ it)
                         (VP was
                             (ADJP-PRD blue)))))
           ball)))
```

3 Punctuation

3.1 Basic guidelines

In this corpus, each unit of text is enclosed in a top level of unlabeled brackets (which generally come out as TOP in the output of the tgrep program). Formerly, top-level punctuation (i.e., initial and final punctuation) could be attached to these top-level brackets. However, in this release, such punctuation should all be attached one level down (to the highest level of *labeled* brackets), so that there is only one top-level node within the unlabeled brackets.

For the sake of simplicity, these unlabeled outer brackets are usually omitted in examples in this manual, and initial and final punctuation are frequently removed as well. However, in this section they are generally included, for greater clarity and precision.

3.1.1 Mid-sentence punctuation

(braces/parentheses, commas, colons, dashes, quotation marks, semicolons)

1. Paired punctuation

   Paired punctuation marks are siblings of the constituent they surround. This is true even when the opening or closing member of the pair can be viewed as deleted. For instance, the commas that set off a subordinate clause or a relative clause from a main clause are siblings of the SBAR dominating the subordinate clause. Similarly, the commas that set off appositives are siblings of the appositive phrase and children of the NP dominating the entire apposition structure.

   Commas setting off appositive phrase:

```
( (S (NP-SBJ (NP John)
             ,
             (NP my brother)
             ,)
     (VP left)
     .))
```

   Commas setting off a parenthetical S:

```
( (S-1 (PP-TMP For
               (NP (NP the rest)
                   (PP of
                       (NP 1989))))
       (PRN ,
            (S (NP-SBJ Mr. Hagen)
               (VP said
                   (SBAR 0
                         (S *T*-1))))
            ,)
       (NP-SBJ (NP Conrail 's)
               traffic and revenue)
       ``
       (VP will
           (VP reflect
               (NP the sluggish economy)))
```

```
      .))
```

Commas setting off a subordinate clause:

```
( (S (SBAR-ADV If
               (S (NP-SBJ-1 the judge)
                  (VP is
                      (VP impeached
                          (NP *-1)))))
     ,
     (SBAR-ADV as
               (S (NP-SBJ-2 *)
                  (VP is
                      (VP thought
                          (S (NP-SBJ *-2)
                             (ADJP-PRD likely))))))
     ,
     (NP-SBJ-3 he)
     (VP will
         (VP be
             (VP removed
                 (NP *-3)
                 (PP-DIR from
                         (NP office))
                 (ADVP-TMP immediately))))
     .))
```

Braces, parentheses, dashes.

Dashes may appear as standard double hyphens (“`--`”) or as single hyphens (“`-`”).

In order to distinguish annotation brackets from brackets that were part of the original text, original brackets are shown with codes:

parentheses () are indicated with `-LRB-` (for Left Round Bracket) and `-RRB-` (for Right Round Bracket).

braces {} are indicated with `-LCB-` and `-RCB-` (for Left/Right Curly Bracket).

brackets [] are indicated with `-LSB-` and `-RSB-` (for Left/Right Square Bracket), but note that the only square brackets in all published versions of the WSJ corpus are right brackets that appear to act as exclamation points.

Most things set off by parentheses or dashes are labeled PRN. Annotator intuition of the interruptiveness of the potential PRN is also a deciding factor in whether or not a word or phrase is labeled PRN; one of the characteristics annotators look for is a compressed pitch range when the sentence is read aloud.

Arguments should never be labeled PRN, though on occasion they have been, due to overzealous annotating. In particular, braces frequently indicate rewording of the quote, so they should almost never get PRN, though occasionally a PRN is appropriate.

Thus, *John (my brother) left*, should be bracketed:

```
( (S (NP-SBJ (NP John)
             (PRN -LRB-
```

```
                  (NP my brother)
                  -RRB-))
     (VP left)
     .))
```

and *"I think {the president} is foolish."* should be bracketed:

```
( (S ``
     (NP-SBJ I)
     (VP think
         (SBAR 0
               (S -LCB-
                  (NP-SBJ the president)
                  -RCB-
                  (VP is
                      (ADJP-PRD foolish)))))
     .
     ''))
```

Quotation marks. These should go outside whatever they surround whenever possible, and when it isn't possible they just get yanked around by whatever is inside them. They are at the very bottom of the pecking order.

In this example the second of the pair of quotes gets yanked to S-level by the comma:

```
( (SINV (S-2 (NP-SBJ Japanese agencies)
             (VP (VP do
                     (NP business)
                     (PP with
                         (NP (NP rival clients)
                             (PP-LOC in
                                     (NP the same industry)))))
                 ,
                 (NP-ADV (NP a practice)
                         ``
                         (SBAR (WHNP-1 that)
                               (S (NP *T* -1)
                                  (VP would
                                      (VP be
                                          (ADJP-PRD unacceptable)
                                          (PP by
                                              (NP traditional Western
                                                  conflict rules)))))))))
        ,
        ''
        (VP says
            (SBAR 0
                  (S *T* -2)))
        (NP-SBJ Roy Warman)
        .))
```

Similarly, when close quotes follow a period, the period takes precedence, i.e. both are attached high:

```
( (S (NP-SBJ Mr. Lorin)
     (VP responded
         , ``
         (INTJ No))
     . ''))


( (S ``
     (NP-SBJ It)
     (VP 's
         not
         (NP-PRD (NP the end)
                 (PP of
                     (NP the world)))
         (SBAR-ADV if
                   (S (NP-SBJ you)
                      (VP shake
                          (NP them)
                          (PRT up)
                          (NP-ADV a little bit)))))
     . ''))
```

Note in the above example how the quotation marks are attached *inside* the highest S.

Exception to the siblings rule for paired punctuation: when an S inside an SBAR is headed by a phrase set off by paired commas, the first comma is placed inside the SBAR, **not** the S:

```
( (S (NP-SBJ I)
     (VP think
         (SBAR that
               ,
               (S (PP-TMP at
                          (NP all times))
                  ,
                  (NP-SBJ Mark)
                  (VP is
                      (ADJP-PRD wrong)))))
     .))
```

However, if the annotator had a strong intuition that the phrase which interjects between the SBAR and the S is a parenthetical remark, he/she may have done the following:

```
(SBAR that
      (PRN ,
           (PP in
               (NP (NP Mr. Letterman 's)
                   words))
           ,)
      (S ...))
```

2. Unpaired punctuation

   Unpaired punctuation intervenes between constituents at the highest possible dividing level:

```
( (S (PP-TMP In
             (NP 1988))
     ,
     (NP-SBJ-1 Dallas-based Sterling)
     (VP protested
         (NP (NP a similar decision)
             (PP by
                 (NP NASA))
             (VP involving
                 (NP the same contract)))
         ,
         (S-ADV (NP-SBJ *-1)
                (VP claiming
                    (SBAR 0
                          (S (NP-SBJ it)
                             (VP had
                                 (VP submitted
                                     (NP the lowest bid))))))))
     .))
```

Colons:

In the sentence *He founded his own company: Crumbly Crackers, Crumbly Crackers* is in apposition to *his own company*. The entire sequence *his own company: Crumbly Crackers* should hence be represented as an adjunction structure, and the colon should be a child of the NP dominating the entire sequence:

```
(NP (NP his own company)
    :
    (NP Crumbly Crackers))
```

Colons in more colorful environments have received various treatments, some of which are shown below:

as sentence-ending punctuation:

```
( (S (NP-SBJ (NP Variation)
             (PP of
                 (NP temperature)))
     (VP gives
         (NP the following results))
     :))
( (S (NP-SBJ (NP an increase)
             (PP in
                 (NP temperature)))
     (VP leads
         (PP-DIR to
                 (NP (NP an increase)
                     (PP of
                         (NP the critical passivation current density)))))
```

```
        .))
```

with pseudo-attach:

```
( (S (NP-SBJ (NP 2 books)
             (NP *ICH*-1))
     (VP are
         (VP recommended
             (PP-MNR with
                     (NP gusto))
             :
             (NP-1 (NP-TTL Harriet the Spy)
                   and
                   (NP-TTL Charlotte's Web))))
     .))
```

3.1.2 Final punctuation

1. Final punctuation

   Final punctuation as a rule is a child of the highest level of structure. Thus, *This is John, my brother*, should be bracketed:

```
( (S (NP-SBJ This)
     (VP is
         (NP-PRD (NP John)
                 ,
                 (NP my brother)))
     .))
```

2. Abbreviation sharing period at end of sentence

   When an abbreviation is the last word in a sentence and its period also serves as the sentence-ending period, the period is attached high rather than to the abbreviated word. If the programs that tokenize the text leave the period attached to the abbreviation, the period is manually split from the abbreviation so that the sentence has final punctuation.

```
( (S (NP-SBJ-1 UAL)
     (VP remains
         (ADJP-PRD obligated
                   (S (NP-SBJ *-1)
                      (VP to
                          (VP pay
                              (NP (NP (QP $ 26.7 million) *U*)
                                  (PP in
                                      (NP fees)))
                              (PP-DTV to
                                      (NP Salomon Brothers Inc)))))))
     .))
```

3.2 Miscellaneous

3.2.1 Symbols functioning as words

Hyphens and other symbols functioning as words are annotated according to their lexical function. Hyphens, for example, are labeled PP when they act prepositionally:

```
(PP (PP from
        (NP 12))
    (PP -
        (NP 15)))
```

3.2.2 Mathematical symbols

Mathematical symbols such as +, -, x, =, etc., are annotated according to their lexical function. For example, "=" is labeled a VP:

```
(S (NP-SBJ x)
   (VP =
       (NP 3)))
```

3.2.3 Ellipses

Three dots (...), usually indicating ellipsis, are treated as a single unit of medial punctuation. Four dots, usually indicating ellipsis at the end of a sentence, are treated as a three-dot ellipsis followed by a period.

```
( (S (NP-SBJ-2 (NP the leadership)
               (PP of
                   (NP the Socialist Unity Party)))
     ,
     (S-ADV (NP-SBJ *-2)
            (VP being
                (ADJP-PRD sensitive
                          (PP to
                              (NP the demands of the time)))))
     ,
     ...
     (VP will
         (VP find
             (NP (NP solutions)
                 (PP to
                     (NP (NP complicated problems)
                         (SBAR (WHNP-1 0)
                               (S (NP-SBJ the German Democratic Republic)
                                  (VP encountered
                                      (NP *T*-1)))))))))
     ... .))
```

4 Null Elements

4.1 The building blocks

4.1.1 Inventory

The inventory of null elements is

T (trace of A$'$-movement, including parasitic gaps)

(NP *) (arbitrary PRO, controlled PRO, and trace of A-movement)

0 (null complementizer, including null *wh*-operator)

U (unit)

? (placeholder for ellipsed material)

NOT (anti-placeholder in template gapping)

RNR (pseudo-attach: right node raising)

ICH (pseudo-attach: interpret constituent here)

EXP (pseudo-attach: expletive)

PPA (pseudo-attach: permanent predictable ambiguity)

But see also section 4.8.7 for list of illegal null elements that may also appear.

Note that while most null elements contain *'s, they are not the only asterisks in the texts; there are also a few naturally-occurring footnote markers. In the "combined" files, null elements are tagged -NONE-, while footnote markers are generally tagged SYM and preceded by a backslash.

4.1.2 Indexing

Indices are used only when they can be used to indicate a relationship that would otherwise not be unambiguously retrievable from the bracketing. Indices are used to express such relationships as coreference (as in the case of controlled PRO or pragmatic coreference for arbitrary PRO), binding (as in the case of *wh*-movement), or close association (as in the case of *it*-extraposition). These relationships are shown only when some type of null element is involved, and only when the relationship is intrasentential. One null element may be associated with another, as in the case of the null *wh*-operator. Coreference relations between overt pronouns and their antecedents are not annotated.

The identity index. In principle, each bracket within the topmost S is understood to have a unique index (an "identity index"), which in practice is used only when that constituent is coreferent with or otherwise closely associated with some null element in the sentence (or when it's acting in a gapping "template"). The brackets surrounding null elements are also understood to be associated with a unique identity index. Identity indices appear only on the bracket label, as in (NP-1 Kris), (WHNP-2 which dog), (SBAR-24 who offered to take me for a swim), etc.

The actual numbering of the identity indices is arbitrary; i.e. the constituents are not necessarily numbered sequentially within the sentence, and a given sentence may contain brackets with the identity indices -1, -2, -5, and -1978. Note also that in rare cases, a bracket may have an identity index shown when there is no corresponding null element or template-gapping constituent.

The reference index. In most cases, a null element will be suffixed with an integer (the "reference index") that matches the identity index on the bracket label of some other constituent. Note that the reference index on the null element takes the form of a dash-number on the null element itself, and not on the bracket label, as in (NP *-1), (NP *T*-2), and (SBAR *ICH*-24). If the null element in turn refers to or is associated with a third element, it will bear its own identity index, along the lines of (NP-1 *T*-2).

```
(S (NP-SBJ-1 he)
   (VP was
       (VP accused
           (NP-3 *-1)
           (PP-CLR of
                   (S-NOM (NP-SBJ *-3)
                          (VP (VP conducting
                                  (NP illegal business))
                              and
                              (VP possessing
                                  (NP illegal materials))))))))
```

```
(S (NP-SBJ (NP It)
           (S *EXP*-2))
   (VP 's
       (ADJP-PRD (ADJP easier)
                 (SBAR *ICH*-1))
       (S-2 (NP-SBJ *)
            (VP to
                (VP get
                    (ADJP-PRD worse))))
       (SBAR-1 than
               (FRAG (ADJP-PRD better)))
       (PP-LOC in
               (NP this game))))
```

(Note that while this indexing system appears complex, the annotation procedure is actually quite simple, as it is accomplished with simple mouse drags that make attention to the gritty details of coindexation unnecessary.)

Indices are also used for the various kinds of pseudo-attach (described in section 5 [Pseudo-Attach]) and for template gapping (described in section 7 [Coordination]).

4.1.3 Other tags

Null elements may bear additional function tags, as described in section 2 [Notation]. For instance, the grammatical function of extracted *wh*-phrases is noted on the *wh*-trace and not on the *wh*-element itself.

```
(SBARQ (WHADVP-439 Where)
       (SQ did
           (NP-SBJ you)
           (VP put
               (NP the book)
               (ADVP-PUT *T*-439)))
       ?)
```

4.2 *T* (trace of A′ movement)

Although the use of *T* corresponds loosely to A′ movement, the match is not precise (e.g. it includes "parasitic gaps"). *T* can also be seen as marking the interpretation location of certain constituents that are not in their usual argument position.

The trace *T* always bears a referential index that corresponds to the identity index of some other constituent in the sentence (moved *wh*-word, topicalized NP or ADVP, etc.).

4.2.1 *Wh*-questions

Wh-moved noun phrases are labeled WHNP and put inside SBARQ. They bear an identity index that matches the reference index on the *T* in the position of the gap. Constituents other than NP are labeled WHxx (WHADVP, WHPP, or WHADJP, as appropriate), placed under SBARQ, and coindexed with the *T* in the position of the gap. The same procedure holds for both arguments and adjuncts.

- NP arguments

```
(SBARQ (WHNP-1 what)
       (SQ are
           (NP-SBJ you)
           (VP thinking
               (PP-CLR about
                       (NP *T*-1))))
       ?)


(SBARQ (WHNP-1 (WHNP Which story)
               (PP about
                   (NP tribbles)))
       (SQ did
           (NP-SBJ you)
           (VP read
               (NP *T*-1)))
       ?)


(SBARQ (WHNP-1 what time)
       (SQ is
           (NP-SBJ it)
           (NP-PRD *T*-1))
       ?)
```

- NP adjunct

```
(SBARQ (WHNP-1 Which day)
       (SQ did
           (NP-SBJ you)
           (VP get
               (ADVP-DIR there)
               (NP-TMP *T*-1)))
       ?)
```

- Non-NP arguments

```
(SBARQ (WHADVP-439 Where)
       (SQ did
           (NP-SBJ you)
           (VP put
               (NP the book)
               (ADVP-PUT *T*-439)))
       ?)


(SBARQ (WHPP-42 On
                (WHNP what))
       (SQ did
           (NP-SBJ you)
           (VP sit
               (PP-LOC-CLR *T*-42)))
       ?)


(SBARQ (WHADJP-54 How cold)
       (SQ is
           (NP-SBJ it)
           (ADJP-PRD *T*-54)
           (ADVP-LOC outside))
       ?)
```

- Non-NP adjuncts

```
(SBARQ (WHADVP-42 How)
       (SQ did
           (NP-SBJ you)
           (VP fix
               (NP the car)
               (ADVP-MNR *T*-42)))
       ?)


(SBARQ (WHADVP-1 Where)
       (SQ did
           (NP-SBJ you)
           (VP meet
               (NP them)
               (ADVP-LOC *T*-1)))
       ?)


(SBARQ (WHADVP-54 Why)
       (SQ did
           (NP-SBJ you)
           (VP jump
               (PP-DIR off
                       (NP the cliff))
```

```
                (ADVP-PRP *T*-54)))
        ?)
```

4.2.2 Relative clauses

Relative clauses are adjoined to the head noun phrase. The relative pronoun is given the appropriate WH-label, put inside the SBAR level, and coindexed with a *T* in the position of the gap. (Note that relative clauses differ from (direct) *wh*-questions in that they contain an SBAR rather than an SBARQ.)

***wh*- and "that" relative clauses.** Relative clauses introduced by *that* are annotated just as relative clauses introduced by a *wh*-word: *that* is given the appropriate WH-label, put inside an SBAR level, and coindexed with the *T* in the position of the gap.

- NP trace

```
(NP (NP answers)
    (SBAR (WHNP-6 that/which)
          (S (NP-SBJ-3 we)
             (VP 'd
                 (VP like
                     (S (NP-SBJ *-3)
                        (VP to
                            (VP have
                                (NP *T*-6)))))))))
```

- ADVP trace

```
(NP (NP the place)
    (SBAR (WHADVP-2 that/where)
          (S (NP-SBJ I)
             (VP put
                 (NP the book)
                 (ADVP-PUT *T*-2)))))
```

Zero relatives. Relative clauses introduced by a null complementizer are annotated in a similar fashion, this time with a null complementizer '0' inside SBAR labeled with the appropriate *wh*-category and coindexed with a *T* in the position of the gap.

- NP trace

```
(NP (NP answers)
    (SBAR (WHNP-3 0)
          (S (NP-SBJ-4 we)
             (VP 'd
                 (VP like
                     (S (NP-SBJ *-4)
                        (VP to
                            (VP have
                                (NP *T*-3)))))))))
```

- ADVP trace

```
(NP (NP the place)
    (SBAR (WHADVP-2 0)
          (S (NP-SBJ I)
             (VP put
                 (NP the book)
                 (ADVP-PUT *T*-2)))))
```

Infinitival relatives. See section 14 [Infinitives] for more information.

- trace as object

```
(NP (NP a movie)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP see
                     (NP *T*-1))))))
```

- trace as subject

```
(NP (NP bloodhounds)
    (SBAR (WHNP-4 0)
          (S (NP-SBJ *T*-4)
             (VP to
                 (VP trail
                     (NP the assassins))))))
```

- trace as adjunct

```
(NP (NP time)
    (SBAR (WHADVP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP go
                     (ADVP-TMP *T*-1))))))
```

4.2.3 Fronted elements

Fronted elements are placed inside the top clause level (e.g. S, SINV, SQ, SBAR). (Only certain fronted elements are tagged -TPC: (i) constituents associated with a *T* in the position of the gap and (ii) left-dislocated constituents (those associated with a resumptive pronoun in the position of the gap).) (See section 1 [Overview of Basic Clause Structure] for more details on the treatment of fronted elements.)

Arguments.

Fronted argument noun phrases are coindexed with a *T* in the position of the gap:

```
(S (NP-TPC-3 This)
   (NP-SBJ every man)
   (VP contains
       (NP *T*-3)
       (PP-LOC-CLR within
                   (NP him))))
```

```
(S (NP-TPC-4 Our dull unsystematic youth)
   (NP-SBJ we)
   (VP let
       (S (NP-SBJ *T*-4)
          (VP stray
              (PP-DIR into
                      (NP philanthropy))))))
```

If the fronted argument is an instance of left-dislocation (i.e, associated with a resumptive pronoun), there is no coindexation between the fronted argument and the pronoun:

```
(S (NP-TPC John)
   ,
   (NP-SBJ I)
   (VP like
       (NP him)
       (NP-ADV a lot)))
```

Other fronted arguments (such as the main VP, a predicate, the locative complement of *put*, etc.) are also tagged -TPC, and their identity index matches the reference index on the *T* inserted in the position of the gap.

```
(S (ADVP-PUT-TPC-1 There)
   ,
   (NP-SBJ I)
   (VP put
       (NP the book)
       (ADVP-PUT *T*-1)))
```

```
(S (SBAR-ADV (VP-TPC-2 Shout
                       (PP-CLR at
                               (NP Eichmann)))
             though
             (S (NP-SBJ he)
                (VP might
                    (VP *T*-2))))
   the prosecutor could not establish...)
```

```
(S (SBAR-ADV (ADJP-PRD-TPC-5 Wrong)
             though
             (S (NP-SBJ the policy)
                (VP may
```

```
                    (VP be
                        (ADJP-PRD *T*-5)))))
     it at least works pretty often.)


  (S (S (NP-SBJ-1 we)
        (VP hope
            (S (NP-SBJ *-1)
               (VP to
                   (VP have
                       (NP a million dollars)
                       (NP-TMP someday))))))
     and
     (S (VP-TPC-6 have
                  (NP it)
                  (SBAR-ADV *ICH*-2))
        (NP-SBJ we)
        (VP may
            (VP *T*-6)
            ,
            (SBAR-ADV-2 if
                        (S (NP-SBJ we)
                           (VP get
                               (ADJP-PRD lucky)))))))
```

Quotations that precede a verb of saying are treated as fronted arguments: they leave a *T* and receive the -TPC tag. (See section 1 [Overview of Basic Clause Structure] for more details on the treatment of quotations.)

```
( (S ``
     (S-TPC-1 (NP-SBJ We)
              (VP will
                  (VP win)))
     ,
     ''
     (NP-SBJ Mary)
     (VP said
         (S *T*-1))
     .))
```

Note that any constituent tagged -CLR is considered an argument for these purposes: it leaves a *T* and receives the -TPC tag if fronted.

```
  (S (PP-CLR-TPC-5 With
                   (NP final exams))
     ,
     (NP-SBJ I)
     (VP associate
         (NP blood , sweat , and tears)
         (PP-CLR *T*-5)))
```

Adjuncts.

Fronted adjuncts are not associated with a *T* when they have not left the clause in which they originate, since in this case their relation to the clause is still clear.

```
(S (ADVP-MNR Carefully)
   ,
   (NP-SBJ I)
   (VP dropped
       (NP the feathers)))
```

```
(S (NP-TMP Yesterday)
   (NP-SBJ-1 a child)
   (VP came
       (ADVP-DIR out)
       (S-PRP (NP-SBJ *-1)
              (VP to
                  (VP wonder)))))
```

```
(SBARQ (NP-TMP Yesterday)
       ,
       (WHNP-2 what)
       (SQ did
           (NP-SBJ we)
           (VP decide
               (NP *T*-2)
               (PP for
                   (NP this one))))
       ?)
```

However, adjuncts that originate in a lower clause are associated with a *T* in the position of the gap.

```
(S (SBAR-PRP-TPC-9 Because
                   (S (NP-SBJ I)
                      (VP 'm
                          (NP-PRD such a bad boy))))
   (NP-SBJ I)
   (VP think
       (SBAR 0
             (S (NP-SBJ I)
                (VP wo n't
                    (VP get
                        (NP a lollipop)
                        (SBAR-PRP *T*-9)))))))
```

In cases where it is ambiguous whether the adjunct originates in a lower clause or in the matrix clause, the adjunct is analyzed as originating from the matrix clause and NOT bracketed -TPC.

4.2.4 Tough movement

The null element is coindexed to a null *wh*-phrase. See section 14 [Infinitives] for more information.

```
(S (NP-SBJ Cars)
   (VP are
```

```
        (ADJP-PRD tough
                  (SBAR (WHNP-7 0)
                        (S (NP-SBJ *)
                           (VP to
                               (VP pay
                                   (PP-CLR for
                                           (NP *T*-7)))))))))
```

4.2.5 Parasitic gaps

A coindexed *T* is put in the parasitic gap as well as in the position of the original gap.

```
(SBARQ (WHNP-1 Which papers)
       (SQ did
           (NP-SBJ-2 you)
           (VP file
               (NP *T*-1)
               (PP without
                   (S-NOM (NP-SBJ *-2)
                          (VP reading (NP *T*-1))))))
       ?)
```

4.3 * (trace of NP movement, controlled PRO, arbitrary PRO)

Because it corresponds to the trace of NP movement, controlled PRO, or arbitrary PRO, * always appears within NP: (NP *).

4.3.1 Indexing

(NP *) bears a reference index whenever it is fairly clear what nominal it is controlled by, corresponding roughly to controlled PRO and the passive trace. However, indexing also reflects pragmatic coreference in addition to syntactic relations, within limits described below and in section 4.8.8.

Unlike *T*, * may appear without a reference index. Unindexed (NP *) corresponds roughly to arbitrary PRO (and passive traces appearing in Reduced Relative Clauses—see section 4.8.2).

In cases of strings of coindexed null elements, the null is coindexed to the most local NP, as with with passives under raising predicates or in the following example:

```
(S (NP-SBJ-1 he)
   (VP was
       (VP accused
           (NP-3 *-1)
           (PP-CLR of
                   (S-NOM (NP-SBJ *-3)
                          (VP (VP conducting
                                  (NP illegal business))
                              and
                              (VP possessing
                                  (NP illegal materials))))))))
```

When several NPs are adjoined, the indexing should be from the highest NP:

```
(S (NP-SBJ-28 (NP Arthur A. Hatch)
              ,
              (NP 59)
              ,)
   (VP was
       (VP named
           (S (NP-SBJ *-28)
              (NP-PRD (NP executive vice president)
                      (PP of
                          (NP the company)))))))
```

4.3.2 Passives

Object of verb. The trace (NP *) is put after the passive verb and coindexed with the constituent in subject position.

```
(S (NP-SBJ-1 John)
   (VP was
       (VP hit
           (NP *-1)
           (PP by
               (NP-LGS a ball)))))
```

Note that the * may come before or after a PRT (particle). There is no policy governing this and either order is possible, though it is somewhat more likely for the PRT to come second:

```
(S (NP-SBJ-1 Arthur)
   (VP was
       (VP picked
           (NP *-1)
           (PRT up)
           (PP by
               (NP-LGS aliens)))))
```

```
(S (NP-SBJ-1 Arthur)
   (VP was
       (VP picked
           (PRT up)
           (NP *-1)
           (PP by
               (NP-LGS aliens)))))
```

Object of preposition. The null may be put after prepositions as required by the construction.

```
(S (NP-SBJ-2 (NP kid 's) cars)
   (VP are
       (ADVP-TMP often)
       (VP paid
           (PP-CLR for
```

```
                (NP *-2))
        (PP by
            (NP-LGS their parents)))))
```

In reduced relative clause. See section 4.8.2 for more details on the treatment of reduced relative clauses. The passive verb (or preposition, as the case may be) in a reduced relative clause is also followed by a (NP *). The passive trace in a reduced relative is not coindexed to the NP preceding it. This reflects an understanding of the relationship between the NP and reduced relative as post-modification rather than predication.

```
(NP (NP an agreement)
    (VP signed
        (NP *)
        (PP by
            (NP-LGS everyone))))
```

```
(NP (NP a government service)
    (VP paid
        (PP-CLR for
                (NP *))
        (PP by
            (NP-LGS everyone))))
```

In some cases (NP *) may function as the subject of a small clause within the reduced relative:

```
(NP (NP an elephant)
    (VP called
        (S (NP-SBJ *)
           (NP-PRD Dumbo))))
```

4.3.3 Subjects of participial clauses and gerunds

The null subject (NP *) of a participial clause or gerund is coindexed with another constituent in the sentence if it is clear to the annotator that the two are coreferent. No attempt is made to conform to the standard Binding Theory of GB or any other such formal approach to coreference. See section 13 [Gerunds and Participles] for more details on the annotation of participial clauses and gerunds.

- VP complements

```
(S (NP-SBJ-1 I)
   (VP stopped
       (S (NP-SBJ *-1)
          (VP eating
              (NP chocolate)))
       (PP-PRP for
               (NP Lent))))
```

```
(S But
   (NP-SBJ-1 I)
```

```
    (VP liked
        (S (NP-SBJ *-1)
           (VP eating
               (NP chocolate)))
        (ADVP-TMP before)))


(S (NP-SBJ-1 he)
   (VP was
       (VP accused
           (NP-3 *-1)
           (PP-CLR of
                   (S-NOM (NP-SBJ *-3)
                          (VP (VP conducting
                                  (NP illegal business))
                              and
                              (VP possessing
                                  (NP illegal materials))))))))
```

- Adverbials

```
(S (NP-SBJ-1 She)
   (VP left
       ,
       (S-ADV (NP-SBJ-2 *-1)
              (VP offended
                  (NP *-2)
                  (PP by (NP-LGS their remarks))))))


(S (NP-SBJ-1 Time)
   (VP eluded
       (NP Paramount)
       (PP-MNR by
               (S-NOM (NP-SBJ *-1)
                      (VP acquiring
                          (NP Warner Communications Inc))))))


(S (NP-SBJ-1 I)
   (VP fell
       (S-CLR (NP-SBJ *-1)
              (ADJP-PRD asleep))
       (PP-LOC on
               (NP the lobby floor))))


(S (NP-SBJ-1 Borough Presidents)
   ,
   (SBAR-ADV while
             (S (NP-SBJ *-1)
                (VP retaining
                    (NP (NP membership)
```

```
                        (PP-LOC in
                                (NP (NP the Board)
                                    (PP of
                                        (NP Estimate)))))))
    ,
    (VP lose
        (NP their housekeeping functions)))
```

- Subjects

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Eating
                  (NP chocolate)))
   (VP is
       (ADJP-PRD good
                 (PP for
                     (NP you)))))
```

- Without coindexation.

  In the case where there is no good candidate for coreference within the sentence, (NP *) remains without an index.

```
(S (NP-SBJ A Texas legislator)
   (VP proposes
       (S (NP-SBJ *)
          (VP color-coding
              (NP (NP (NP drivers ')
                      licenses)
                  (PP of
                      (NP some drug offenders)))))))
```

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Taking
                  (NP Iwo Jima)))
   (VP was
       (NP-PRD no easy feat)))
```

  Note also that the null subject of a gerund that is coordinated with one or more NPs is usually not coindexed.

```
(S (NP-SBJ I)
   (VP like
       (NP (NP medals)
           ,
           (NP cheering crowds)
           ,
           and
           (S-NOM (NP-SBJ *)
                  (VP swimming
                      (NP the backstroke))))))
```

4.3.4 Subjects of infinitival clauses

With coindexation

1. VP complement clauses.

   Note that from the perspective of the annotator, it is not necessary to distinguish between Raising and Control structures, etc. In each case, the annotator simply coindexes the empty subject of the infinitival with whatever lexical NP it is associated with.

   (a) "Raising" constructions.

```
(S (NP-SBJ-3 Everyone)
   (VP seems
       (S (NP-SBJ *-3)
          (VP to
              (VP dislike
                  (NP Drew Barrymore))))))
```

   (b) "Object control" constructions.

```
(S (NP-SBJ Ford)
   (VP persuaded
       (NP-1 Zaphod)
       (S (NP-SBJ *-1)
          (VP to
              (VP run
                  (PP-CLR for
                          (NP president)))))))
```

   (c) "Subject control" constructions.

```
(S (NP-SBJ-1 Zaphod)
   (VP promised
       (NP Ford)
       (S (NP-SBJ *-1)
          (VP to
              (VP run
                  (PP-CLR for
                          (NP president)))))))
```

   (d) Semi-auxiliaries.

   Semi-auxiliaries occur in constructions with infinitival *to*, (e.g, *supposed to, ought to, have to*). They are annotated with full infinitival structure and have a (NP-SBJ *) subject, coindexed as appropriate.

```
(S (PP Of (NP course))
   ,
   (NP-SBJ-1 regulators)
   (VP would
       (VP have
           (S (NP-SBJ *-1)
              (VP to
                  (VP approve
                      (NP (NP Columbia 's)
```

```
                                reorganization)))))))
```

about to is also treated as a semi-auxiliary in official policy, though some variation exists. The following bracketings are likely.

- Official Policy:

```
(S (NP-SBJ-2 interest rates)
   (VP are
       (VP about
           (S (NP-SBJ *-2)
              (VP to
                  (VP nosedive))))))
```

- Common alternate:

```
(S (NP-SBJ-1 the pollen)
   (VP is
       (ADJP-PRD about
                 (S (NP-SBJ *-1)
                    (VP to
                        (VP develop))))))
```

2. ADJP complement clauses.

   Null element is coindexed with the matrix subject, where appropriate.

   (a) "Raising" constructions.

```
(S (NP-SBJ-4 This climb)
   (VP is
       (ADJP-PRD likely
                 (S (NP-SBJ *-4)
                    (VP to
                        (VP be
                            (ADJP-PRD difficult)))))))
```

   (b) "Control" constructions.

```
(S (NP-SBJ-1 Zaphod)
   (VP is
       (ADJP-PRD ready
                 (S (NP-SBJ *-1)
                    (VP to
                        (VP eat
                            (NP the steak)))))))
```

3. Adverbials.

   Here are a few common adverbials infinitives. This is not an exhaustive list.

   (a) purpose clause

```
(S (NP-SBJ-1 The public)
   (VP did n't
       (VP come
           (PP-DIR to
                   (NP the market))
```

```
            (S-PRP (NP-SBJ *-1)
                   (VP to
                       (VP play
                           (NP a game)))))))
```

(b) semi-complement clause

```
(S (NP-SBJ-1 Skilled ringers)
   (VP use
       (NP their wrists)
       (S-CLR (NP-SBJ *-1)
              (VP to
                  (VP advance or retard
                      (NP the next swing))))))
```

(c) resultative clause

```
(S (NP-SBJ-1 (NP London 's)
             Financial Times 100-share index)
   (VP shed
       (NP 40.4 points)
       (S-ADV (NP-SBJ *-1)
              (VP to
                  (VP finish
                      (PP-CLR at
                              (NP 2149.3)))))))
```

Without coindexation. The following types of (NP *) subject are not coindexed: subject of infinitive inside NP, imperative subject, subject in “tough-movement” construction. Aside from these rules, there are some cases in which coindexation is much less likely than normal, described in section 4.8.8.

1. Infinitives inside NPs.

   These include complement clauses within NP and the subjects of infinitival relative clauses,

```
(NP (NP John 's)
    decision
    (S (NP-SBJ *)
       (VP to
           (VP leave))))
```

```
(S (NP-SBJ I)
   (VP made
       (NP a decision
           (S (NP-SBJ *)
              (VP to
                  (VP leave))))))
```

```
(NP (NP a manual)
    (SBAR (WHNP-1 0)
```

```
            (S (NP-SBJ *)
               (VP to
                   (VP write
                       (NP *T*-1))))))
```

Note, however, that the *T* subject of an infinitival relative is coindexed as appropriate.

```
(NP (NP bloodhounds)
    (SBAR (WHNP-4 0)
          (S (NP-SBJ *T*-4)
             (VP to
                 (VP trail
                     (NP the assassins))))))
```

2. Imperative subjects.

```
(S (NP-VOC Kris)
   ,
   (NP-SBJ *)
   (VP go
       (ADVP-DIR home))
   !)
```

3. Subjects in "tough"-movement constructions.

```
(S (PP for (NP Zaphod))
   ,
   (NP-SBJ that steak)
   (VP is
       (ADJP-PRD ready
                 (SBAR (WHNP-1 0)
                       (S (NP-SBJ *)
                          (VP to
                              (VP eat
                                  (NP *T*-1))))))))
```

4.3.5 Subjects of *as*- and *than*-clauses

A (NP *) is used as a "placeholder" subject in clauses introduced by *than* or *as* that lack an overt subject. This type of structure may also be annotated using *?* or FRAG, though both are rare. See then end of section 4.6.2 (page 87) for more on these possibilities.

```
(S But
   (NP-SBJ there)
   (VP may
       (VP be
           (NP-PRD (NP less)
                   (ADVP-LOC there)
                   (SBAR than
```

```
                          (S (NP-SBJ *)
                             (VP meets
                                 (NP the eye)))))))
```

```
(NP (NP as little)
    (SBAR as
          (S (NP-SBJ *)
             (VP is
                 (ADJP-PRD consistent ...)))))
```

```
(S (NP-SBJ Primerica)
   ,
   (SBAR-ADV as
             (S (NP-SBJ *)
                (VP expected)))
   ,
   (ADVP also)
   (VP acquired ...))
```

```
(NP (NP Items)
    (VP listed
        (NP-1 *)
        (PP-CLR as
                (S-NOM (NP-SBJ *-1)
                       (VP being
                           (PP-PRD in
                                   (NP short supply)))))))
```

```
(S (NP-SBJ-1 She)
   (VP was
       (VP quoted
           (NP-2 *-1)
           (PP-CLR as
                   (S-NOM (NP-SBJ *-2)
                          (VP saying ...))))))
```

4.4 0 (null complementizer)

`0` is used inside SBAR only if there is no overt *wh*-element or *that*: `(SBAR 0)` or `(SBAR (WHxx 0))`.

4.4.1 Subordinator for tensed complement clauses

The null complementizer introduces most tensed complement clauses.

- With complement of ADJP.

```
(S (NP-SBJ I)
   (VP 'm
       (ADJP-PRD sure
```

```
            (SBAR 0
                  (S (NP-SBJ he)
                     (VP 'll
                         (VP be
                             (ADVP-LOC-PRD here)
                             (NP-TMP any minute)))))))
```

- With complement of VP.

```
(S (NP-SBJ I)
   (VP believe
       (SBAR 0
             (S (NP-SBJ you)
                (VP are
                    (ADJP-PRD *?*))))))
```

- With complement of NP.

```
(S (NP-SBJ he)
   (VP wrote
       (SBAR that
             (S (NP-SBJ he)
                (VP had
                    (VP given (PRT up)
                        (NP hope
                            (SBAR 0
                                  (S (NP-SBJ they)
                                     (VP would
                                         (ADVP-TMP ever)
                                         (VP agree
                                             (PP-CLR on
                                                     (NP anything)))))))))))))
```

```
(S (PP in
       (NP the event
           (SBAR 0
                 (S (NP-SBJ Congress)
                    (VP does
                        (VP provide
                            (NP (NP this increase)
                                (PP in (NP federal funds)))))))))
   ,
   (NP-SBJ the State Board)
   (VP should ...))
```

4.4.2 Zero relative clauses

- The null complementizer is labeled (WHNP 0) if it corresponds to *who, which, that,* etc.

```
(NP (NP the bird)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ I)
             (VP saw
                 (NP *T*-1)))))
```

- The null complementizer is labeled (WHADVP 0) if it corresponds to *where, why, when, how,* etc.

```
(NP (NP the place)
    (SBAR (WHADVP-2 0)
          (S (NP-SBJ I)
             (VP put
                 (NP the book)
                 (ADVP-PUT *T*-2)))))
```

```
(NP (NP the reason)
    (SBAR (WHADVP-3 0)
          (S (NP-SBJ he)
             (VP came
                 (ADVP-PRP *T*-3)))))
```

4.4.3 Infinitival relative clauses

See section 14 [Infinitives] for more details on Infinitives.

- (WHNP 0) is used for NP objects and subjects.

```
(NP (NP a movie)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP see
                     (NP *T*-1))))))
```

- (WHADVP 0) is used in the case where the missing element can be paraphrased as *in which, at which, for which,* etc.

```
(S (NP-SBJ That)
   (VP 's
       (NP-PRD (NP a good way)
               (SBAR (WHADVP-1 0)
                     (S (NP-SBJ *)
                        (VP to
                            (VP keep
                                (ADJP-PRD warm)
                                (ADVP-MNR *T*-1))))))))
```

```
(S (NP-SBJ There)
   (VP 's
       (NP-PRD (NP no reason)
               (SBAR (WHADVP-1 0)
                     (S (NP-SBJ *)
                        (VP to
                            (VP do
                                (NP it)
                                (NP-MNR that way)
                                (ADVP-PRP *T*-1))))))))
```

- When the infinitival is introduced by *for*, the appropriate form of (WHxx 0) is inserted **before** the *for* in SBAR:

```
(NP (NP a movie)
    (SBAR (WHNP-3 0)
          for
          (S (NP-SBJ us)
             (VP to
                 (VP see
                     (NP *T*-3))))))


(NP (NP a good way)
    (SBAR (WHADVP-4 0)
          for
          (S (NP-SBJ them)
             (VP to
                 (VP do
                     (NP it)
                     (ADVP-MNR *T*-4))))))
```

4.5 *U* (unit)

This element marks the interpreted position of a unit symbol, such as $, # (British pounds), FFr (French francs), C$, US$, HK$, A$, M$, S$, and NZ$. It may also appear after % or even *cents*, when convenient. See section 11 [Modification of NP] for more details on the use of *U*.

4.5.1 Uses

```
(NP C$ 5 *U*)

(NP (QP between $ 5 and $ 15) *U*)
```

After %, *U* is used only as necessary.

```
(NP (QP between 12 % to 13 %)
    *U*)
```

After *cents*, *U* has been used occasionally for certain complex constructions, though such use is not officially sanctioned.

```
(NP (NP earnings)
    (PP of
        (NP (NP (QP between 62 cents and 64 cents)
                *U*)
            (NP-ADV a share))))
```

4.5.2 Placement

In general, *U* is placed where the word corresponding to the symbol would appear in the string if the text were read aloud. One notable exception is in certain hyphenated compound adjectives, such as *a $5-a-share increase* (spoken: "A five dollar a share increase"). Here, the bracketing will usually not reflect the spoken order, with *U* placed as the last element in the ADJP:

```
(NP a (ADJP $ 5-a-share *U*)
    increase)
```

Sometimes, this type may lack the *U* entirely.

4.6 *?* (placeholder for ellipsed material)

? is now available in the following great-tasting flavors: (VP *?*), (ADJP-PRD *?*), (PP-PRD *), (NP *?*), (S *?*), (SBAR *?*). These act as placeholders for a missing predicate or piece thereof, especially in comparative constructions and other environments where predicate deletion occurs. Although the missing material represented by *?* is often identical to another constituent in the same sentence, the two are never coindexed. Postmodifiers of the verb (including traces) may be attached under (VP *?*), but not to any other null element, including the other *?* null elements and (VP *T*).

Note that policy for *?* was never finalized, so its use varies to some extent. In general, *?* is used by the annotators as a last resort (short of the FRAG analysis) for the annotation of clauses with "missing" material. Nonetheless, there are certain constructions that are particularly likely to contain *?*:

4.6.1 Comparative deletion

(See section 22 [Comparatives] for more information.)

Complement of *be*. Missing complements of *be* are labeled as appropriate and receive the -PRD tag.

```
(S (NP-SBJ John)
   (VP is
       (ADJP-PRD (ADJP sillier)
                 (SBAR than
                       (S (NP-SBJ I)
                          (VP am
                              (ADJP-PRD *?*)))))))
```

```
(S (NP-SBJ Laos)
   (VP is
       (PP-PRD of
               (NP (NP (ADVP no more)
                       (ADJP purely military)
                       value)
                   (SBAR *ICH*-2)))
       (PP to
           (NP (NP Moscow)
               (NP itself)))
       (SBAR-2 than
               (S (NP-SBJ it)
                  (VP is
                      (PP-PRD *?*)
                      (PP to
                          (NP Washington)))))))
```

Direct object.

```
(S (NP-SBJ the Controller)
   (VP will
       (VP have
           (NP (NP the opportunity)
               (PP for
                   (NP (NP greater usefulness)
                       (PP to
                           (NP good government))
                       (SBAR than
                             (S (NP-SBJ he)
                                (VP has
                                    (NP *?*)
                                    (ADVP-TMP now))))))))))
```

```
(S (NP-SBJ-2 the Fed)
   (VP was
       (VP prepared
           (S (NP-SBJ *-2)
              (VP to
                  (VP provide
                      (NP (NP (ADJP as much) credit)
                          (SBAR as
                                (S (NP-SBJ the markets)
                                   (VP needed
                                       (NP *?*))))))))))))
```

Verb phrase.

```
(S (NP-SBJ (NP The submission)
           (PP of (NP detailed plans)))
   (VP would
```

```
        (VP place
            (NP the issues)
            (PP-LOC-CLR before (NP the court))
            (ADVP-MNR (ADVP more readily)
                      (SBAR than
                            (SINV would
                                  (NP-SBJ (NP discussion)
                                          (PP of
                                              (NP divestiture
                                                  or
                                                  disenfranchisement))
                                          (PP in (NP the abstract)))
                                  (VP *?*)))))))
```

Clausal complement.

```
(S (PP As
       (NP Mayor))
   ,
   (NP-SBJ-1 Mr. Levitt)
   (VP might
       (VP turn
           (PRT out)
           (S (NP-SBJ *-1)
              (VP to
                  (VP be
                      (ADJP-PRD (ADJP more independent)
                                (SBAR than
                                      (S (NP-SBJ (NP some)
                                                 (PP of
                                                     (NP his leading
                                                         supporters)))
                                         (VP would
                                             (VP like
                                                 (S *?*))))))))))))
```

```
(S (NP-SBJ the steel strike)
   (VP lasted
       (ADVP-TMP (ADVP much longer)
                 (SBAR than
                       (S (NP-SBJ he)
                          (VP anticipated
                              (SBAR 0
                                    (S *?*))))))))
```

VP pro-form *do.* In constructions with a VP pro-form, a missing VP may be postulated and shown as the complement of *do*.

```
(S (NP-SBJ Bill)
   (VP eats
```

```
(NP (NP more hotdogs)
    (SBAR than
          (S (NP-SBJ Mary)
             (VP does
                 (VP *?*)))))))
```

4.6.2 Deletion in non-comparatives

VP after missing auxiliary in second conjunct.

```
(S (S (NP-SBJ She)
      (ADVP-TMP rarely)
      (VP sings))
   ,
   so
   (S (NP-SBJ I)
      (VP do n't
          (VP think
              (SBAR 0
                    (S (NP-SBJ she)
                       (VP will
                           (VP *?*
                               (NP-TMP tonight)))))))))
```

```
(S (S (NP-SBJ Robin)
      (VP likes
          (NP ice cream)))
   , and
   (S (NP-SBJ Kim)
      (VP does
          (VP *?*
              ,
              (ADVP too)))))
```

VP after missing auxiliary in subordinate clauses.

```
(S (NP-SBJ *)
   (VP Call
       (S (NP-SBJ it)
          (ADJP-PRD anecdotal))
       (SBAR-ADV if
                 (S (NP-SBJ you)
                    (VP will
                        (VP *?*))))))
```

```
(S (SBAR-ADV As
             (S (NP-SBJ they)
                (VP did
                    (VP *?*
                        (SBAR-TMP (WHADVP-1 when)
```

```
                                    (S (NP-SBJ the Philippines)
                                       (VP was
                                           (NP-PRD a colony)
                                           (ADVP-TMP *T*-1))))))))
   ,
   (NP-SBJ teachers)
   (PP for
       (NP the most part))
   (VP teach
       (PP-MNR in
               (NP English))))
```

```
(S (NP-SBJ they)
   (VP must
       (VP buy
           (NP shares)
           (PP-CLR from
                   (NP sellers))
           (SBAR-TMP (WHADVP-1 when)
                     (S (NP-SBJ-2 (NP no one)
                                  (ADJP else))
                        (VP is
                            (ADJP-PRD willing
                                      (S (NP-SBJ *-2)
                                         (VP to
                                             (VP *?*))))))))))
```

VP missing in *as do*-type constructions

```
(S (NP-SBJ (NP Warner)
           and
           (NP Mr. Azoff))
   (VP declined
       (NP comment)
       ,
       (SBAR-ADV as
                 (SINV did
                       (NP-SBJ MCA)
                       (VP *?*)))))
```

Non-inverted version (*as MCA did*) might also have (VP *?*).

```
(SBAR-ADV as
          (S (NP-SBJ MCA)
             (VP did
                 (VP *?*))))
```

But note that when the *so* predicate pro-form is present, *?* is not used, and the *so* is treated as an adverbial predicate standing in for the understood verbal predicate:

```
(S (NP-SBJ They)
   (ADVP also)
```

```
    (VP did
        (ADVP-PRD so)))

 (S (S (NP The winners)
       (VP had
           (NP fun)))
    ,
    and
    (SINV (ADVP-PRD-TPC-1 so)
          (VP did
              (ADVP-PRD *T*-1))
          (NP-SBJ the losers)))
```

Note, however, that this policy was late in appearance and not always well understood, so the -PRD label may be missing, or replaced by -CLR or -MNR. Also, in inverted sentences, the (ADVP so) may lack the -TPC and accompanying *T*.

With relation to previous sentence. The following two sentences appear in succession (in wsj_2106). The second conjunct of the first sentence contains *?*, indicating ellipsis with respect to material contained in the first conjunct. The second sentence also contains *?*, again indicating ellipsis with respect to the same material. As a rule, we do not indicate intersentential relationships, but here the second instance of *?* is present by virtue of this intersentential relationship, though such relationship is not explicitly recoverable from the annotation.

```
( (S Either
     (S (NP-SBJ one)
        (VP likes
            (NP it)))
     or
     (S (NP-SBJ one)
        (VP does n't
            (VP *?*)))
     .))
( (S (NP-SBJ The typical Glass audience)
     (ADVP certainly)
     (VP does
         (VP *?*))
     .))
```

In subject position. In very rare (about 5) cases, *?* may also appear in subject position, although * is much more likely to be found there.

```
 (S (NP-SBJ We)
    (VP are
        (VP working
            (ADVP (ADVP significantly longer and harder)
                  (SBAR than
                        (S (NP-SBJ *?*)
                           (VP has
                               (VP been
                                   (NP-PRD the case)
```

```
                                    (PP-TMP in
                                            (NP the past))))))))))
```

Other possibilities include FRAG:

```
(S (NP-SBJ-1 we)
   (VP will
       (VP take
           (NP ongoing cost-reduction actions)
           (SBAR-ADV as
                     (FRAG (ADJP necessary))))))
```

and (most common by far) simply calling the *as* or *than* PP:

```
(NP-PRD (NP another day)
        (PP of
            (NP (NP ectoplasmic business)
                (PP as
                    (ADJP usual)))))


(S (NP-SBJ The decline)
   (VP was
       (ADJP-PRD (ADJP even steeper)
                 (PP than
                     (PP-TMP in
                             (NP September))))))
```

4.6.3 Undefined gaps

? is occasionally used to fill a (noun-phrase) gap for which there is no well-established policy:

```
(S (NP-SBJ The plant)
   (VP will
       (VP cost
           (NP (QP about 50 million) Canadian dollars)
           (S-CLR (NP-SBJ *)
                  (VP to
                      (VP build
                          (NP *?*)))))))


(NP (NP a return)
    (ADJP worth
          (S (NP-SBJ *)
             (VP getting
                 (ADJP-PRD excited
                           (PP about
                               (NP *?*)))))))
```

4.7 *NOT* (anti-placeholder in template gapping)

NOT is used in the template gapping procedure, along with "=" and a system of coindexation. See section 7 [Coordination] for more details on the template gapping approach.

Unlike other null elements, correspondence to a *NOT* is shown by an "=" index on the bracket **label**, rather than by a "–" index on the null element itself.

NOT is used very rarely (about 20 times in the entire WSJ corpus) when the "template" and "copy" are not entirely parallel. In principle, it can serve in two roles:

4.7.1 In the "copy"

In the **copy**, *NOT* can be used to indicate that the corresponding constituent in the template is **not** interpreted in the "copy".

```
(S (NP-SBJ the auditor)
   (VP (VP (ADVP-TMP-2 first)
           described
           (NP-3 the old plan)
           (PP-CLR-4 as
                     (ADJP ill conceived))
           (PRN -LRB-
                (SBAR-ADV-5 as
                            (S (NP-SBJ everyone)
                               (ADVP-TMP already)
                               (VP agreed)))
                -RRB-))
       but
       (VP (ADVP-TMP=2 then)
           (NP=3 (NP the new plan)
                 (SBAR (WHNP-1 that)
                       (S (NP-SBJ we)
                          (VP 'd
                              (VP worked
                                  (ADVP-MNR so hard)
                                  (PP-CLR on
                                          (NP *T*-1)))))))
           (PP-CLR=4 as
                     (UCP (PP out (PP to (NP lunch)))
                          and
                          (ADJP totally half-baked)))
           (ADVP=5 *NOT*))))
```

Note that it is very difficult to construct a grammatical example, so (unsurprisingly) there are no actual examples of this construction in the WSJ corpus.

4.7.2 In the "template"

In the **template**, *NOT* appears in the position where the corresponding constituent in the copy is interpreted, when there is no matching constituent already in the template.

```
(S (NP-SBJ (NP The 189 Democrats)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP supported
                        (NP the override)
                        (NP-TMP yesterday)))))
   (VP compare
       (PP-CLR with
               (NP (NP (NP-3 175)
                       (SBAR (WHNP-2 who)
                             (S (NP-SBJ *T*-2)
                                (ADVP initially)
                                (VP backed
                                    (NP the rape-and-incest exemption)
                                    (ADVP-TMP-4 (NP two weeks)
                                                ago)
                                    (PP-5 *NOT*)))))
                   and
                   (NP (NP=3 136)
                       (NP-TMP=4 last year)
                       (PP=5 on
                             (NP a similar vote)))))))
```

More often than not, however, it is assumed that an unindexed constituent at top level of the "copy" is interpreted at highest possible level – usually VP-level of the template. (Note that this rule doesn't work in the above example, so it actually needs the *NOT*.)

So the following examples:

```
(S (NP The teacher)
   (VP (VP gave
           (NP-1 Ignatius)
           (NP-2 only a B)
           (SBAR-3 *NOT*))
       ,
       but
       (VP (NP=1 Bertha)
           (NP=2 an A)
           (SBAR-PRP=3 because
                       (S (NP-SBJ she)
                          (VP wrote
                              (ADVP-MNR so well)))))))


(VP (VP increasing
        (PP-DIR-2 to
                  (NP 2.5 %))
        (PP-TMP-3 in
                  (NP February 1991))
        (ADVP-TMP-4 *NOT*))
    ,
    and
    (VP (PP-DIR=2 to
                  (NP 3 %))
```

```
        (PP-TMP=3 at
                  (NP six-month intervals))
        (ADVP-TMP=4 thereafter)))
```

might more likely be bracketed:

```
(S (NP The teacher)
   (VP (VP gave
           (NP-1 Ignatius)
           (NP-2 only a B))
       ,
       but
       (VP (NP=1 Bertha)
           (NP=2 an A)
           (SBAR-PRP because
                     (S (NP-SBJ she)
                        (VP wrote
                            (ADVP-MNR so well)))))))
```

```
(VP (VP increasing
        (PP-DIR-2 to
                  (NP 2.5 %))
        (PP-TMP-3 in
                  (NP February 1991)))
    ,
    and
    (VP (PP-DIR=2 to
                  (NP 3 %))
        (PP-TMP=3 at
                  (NP six-month intervals))
        (ADVP-TMP thereafter)))
```

4.7.3 Alternatives to *NOT*

Certain constructions that are sometimes analyzed using *NOT* are more likely to be analyzed more simply, usually using PRN or FRAG or just PP.

- complicated *NOT* analysis

```
(S (S (NP-SBJ This gap)
      (ADVP-TMP-2 eventually)
      (VP closes
          (ADVP-MNR-1 *NOT*)))
   ,
   but
   (S (ADVP-TMP=2 *NOT*)
      (ADVP-MNR=1 slowly)))
```

- PP analysis

```
(S (NP-SBJ This gap)
   (ADVP-TMP eventually)
   (VP closes
       ,
       (PP but
           (ADVP-MNR slowly))))
```

- PRN analysis

```
(S (NP-SBJ This gap)
   (ADVP-TMP eventually)
   (VP closes
       (PRN ,
            but
            (ADVP-MNR slowly))))
```

- FRAG analysis

```
(S (S (NP-SBJ This gap)
      (ADVP-TMP eventually)
      (VP closes))
   ,
   but
   (FRAG (ADVP-MNR slowly)))
```

- conjunction? what conjunction?

```
(S (NP-SBJ This gap)
   (ADVP-TMP eventually)
   (VP closes
       ,
       but
       (ADVP-MNR slowly)))
```

4.8 Miscellaneous

4.8.1 Subject-aux inversion with subject extractions

When a question is of the subject, the placement of the null *T* subject determines whether the resulting clause appears to exhibit subject-aux inversion.

In most cases, the (NP-SBJ *T*) is inserted at the beginning of the SQ, before the inflected verb or auxiliary:

```
(SBARQ (WHNP-4 Who)
       (SQ (NP-SBJ *T*-4)
           (VP came
               (PP-DIR to
                       (NP the party))))
       ?)
```

```
(SBARQ (WHNP-1 Who)
       (SQ (NP-SBJ *T*-1)
           (VP will
               (VP come
                   (PP-DIR to
                           (NP the party))))))
```

```
(SBARQ (WHNP-4 Who)
       (SQ (NP-SBJ *T*-4)
           (ADVP-TMP always)
           (VP comes
               (PP-DIR to
                       (NP parties))))
       ?)
```

```
(SBARQ (WHNP-1 Who)
       (SQ (NP-SBJ *T*-1)
           (VP will
               (ADVP-TMP never)
               (VP come
                   (PP-DIR to
                           (NP the party))))))
```

Note, however, that there is some variation (since annotators are accustomed to seeing subject-aux inversion in many questions), and the subject trace is occasionally inserted directly after an initial auxiliary:

```
(SBARQ (WHNP-1 Who)
       (SQ will
           (NP-SBJ *T*-1)
           (VP come
               (PP-DIR to
                       (NP the party)))))
```

```
(SBARQ (WHNP-1 Who)
       (SQ will
           (NP-SBJ *T*-1)
           (ADVP-TMP never)
           (VP come
               (PP-DIR to
                       (NP the party)))))
```

4.8.2 Reduced relatives

Reduced relative clauses are bracketed as follows, with no structure above the VP level:

```
(NP (NP an elephant)
    (VP called
        (S (NP-SBJ *)
           (NP-PRD Dumbo))))
```

It may be that the underlying structure of the reduced relative is as follows:

```
(NP (NP an elephant)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ-2 *T*-1)
             (VP BE
                 (VP called
                     (S (NP-SBJ *-2)
                        (NP-PRD Dumbo)))))))
```

But we do not attempt to reflect this particular understanding of the reduced relative in the annotation. Nonetheless, the bracketing as it is can automatically be transformed into such a structure. Note also that choosing the former style of annotation over the latter necessarily affects coindexation: The passive (NP *) does not bear an index in the former, as the null subject with which it is associated is not present in the annotation.

4.8.3 Attachment to null elements

Barring certain exceptions, a null element never has another constituent attached to it. Thus, the SBAR in the first example below and the PP in the second example are not attached to the trace, but rather to its overt associate NP elsewhere in the sentence.

```
(SBARQ (WHNP-2 (WHNP Who)
               (SBAR *ICH*-3))
       (SQ did
           (NP-SBJ you)
           (VP meet
               (NP *T*-2)
               (SBAR-3 (WHNP-1 that)
                       (S (NP-SBJ *T*-1)
                          (VP wore
                              (NP overalls)))))))


(SINV (ADVP-LOC-PRD-TPC-5 Here)
      (VP are
          (ADVP-LOC-PRD *T*-5))
      (NP-SBJ (NP the pictures)
              (SBAR (WHNP-3 0)
                    (S (NP-SBJ-1 you)
                       (VP wanted
                           (S (NP-SBJ *-1)
                              (VP to
                                  (VP see
                                      (NP *T*-3)))))))
              (PP of
                  (NP (NP that cute dog)
                      (SBAR (WHNP-2 0)
                            (S (NP-SBJ we)
                               (VP met
                                   (NP *T*-2)
                                   (NP-TMP the other day))))))))
```

Note that (VP *?*) (but **not** (VP *T*-1)) is an exception to this rule.

```
(S (S (NP-SBJ She)
      (ADVP-TMP rarely)
      (VP sings))
   ,
   so
   (S (NP-SBJ I)
      (VP do n't
          (VP think
              (SBAR 0
                    (S (NP-SBJ she)
                       (VP will
                           (VP *?*
                               (NP-TMP tonight)))))))))
```

4.8.4 Attachment of null elements

***PPA* (permanent predictable ambiguity).** *PPA*-attach is used to indicate ambiguity of attachment of a trace, if the sentence is truly ambiguous (here "why was the decision made" vs. "why do you think it was made"):

```
(SBARQ (WHADVP-1 Why)
       (SQ do
           (NP-SBJ you)
           (VP think
               (SBAR 0
                     (S (NP-SBJ we)
                        (VP made
                            (NP that decision)
                            (ADVP-PRP *PPA*-2))))
               (ADVP-PRP-2 *T*-1))))
```

Note that such ambiguity is unlikely context, so such examples are rare or nonexistent in the actual corpus.

Shared traces. On the other hand, **shared** traces are handled quite differently. When a trace is interpreted as part of two separate conjuncts, there will be one trace at conjuction level if the element in question is a VP adjunct:

```
(NP (NP a business system)
    (SBAR (WHADVP-1 where)
          (S (NP-SBJ shareholders)
             (VP (VP have
                     (NP few rights))
                 and
                 (VP expect
                     (NP only modest dividends))
                 (ADVP-LOC *T*-1)))))
```

...but two separate traces otherwise (pseudo-attach is not used):

```
(S (NP-SBJ-2 (NP (QP No fewer than 24)
                 country funds))
   (VP have
       (VP been
           (VP (VP launched
                   (NP *-2))
               or
               (VP registered
                   (NP *-2)
                   (PP-CLR with
                           (NP regulators)))
               (NP-TMP this year)))))
```

```
(S (PP-TPC-1 Of (NP the 13 entrants))
   ,
   (S (NP-SBJ (NP 5)
              (PP *T*-1))
      (VP finished))
   and
   (S (NP-SBJ (NP 8)
              (PP *T*-1))
      (VP crashed)))
```

4.8.5 Interpreting the WH label

WHNP, WHADJP, etc. are labels that mark a *wh*-phrase in SBAR that has an associated trace *T* in the position where the *wh*-phrase is interpreted.

Wh-phrases usually contain a *wh*-word, such as *who, whose, which, when, where, how, why, whom, whenever, whatever,* etc. in questions and relative clauses. The WHx label is also used for *that* and 0 (zero) in relative clauses.

The label applies only to *wh*-words that appear in SBAR. It is not used for *in situ wh*-phrases:

```
(S (NP-SBJ You)
   (VP said
       (NP what))
   ?)
```

```
(S (NP-SBJ The butcher)
   (VP gave
       (NP the bone)
       (PP-DTV to
               (NP which dog)))
   ?)
```

However, note that in sentence fragments where the relative position of the *wh*-word may not be clear, the *wh*-word in question may be labeled as such if the annotator has the sense that the wh word is not *in situ*.

```
(FRAG (WHADVP Why)
      not
      ?)
```

```
( (S (NP-SBJ You)
     (VP said
         (SBAR 0
               (S (NP-SBJ John)
                  (VP gave
                      (NP you)
                      (NP something)))))
     .))
( (FRAG (WHNP what)
        ?))
```

In complex *wh*-phrases, *wh*-ness percolates up but not down. Thus, in the following, *from* is labeled WH but *of* and *syntax* are not.

```
  (SBARQ (WHPP-1 From
                 (WHNP (WHNP whose theory)
                       (PP of
                           (NP syntax))))
         (SQ do
             (NP-SBJ you)
             (VP draw
                 (NP this conclusion)
                 (PP-CLR *T*-1)))
         ?)
```

4.8.6 Comparative relatives

Certain comparatives may have been analyzed as relative clauses, as in (b), although the *?* analysis is more common, as in (a):

```
a.  (S (NP-SBJ-2 the Fed)
       (VP was
           (ADJP-PRD prepared
                     (S (NP-SBJ *-2)
                        (VP to
                            (VP provide
                                (NP (NP (ADJP as much) credit)
                                    (SBAR as
                                          (S (NP-SBJ the markets)
                                             (VP needed
                                                 (NP *?*))))))))))))


b.  (S (NP-SBJ-2 the Fed)
       (VP was
           (ADJP-PRD prepared
                     (S (NP-SBJ *-2)
                        (VP to
                            (VP provide
                                (NP (NP (ADJP as much) credit)
                                    (SBAR (WHNP-3 0)
                                          as
```

```
(S (NP-SBJ the markets)
   (VP needed
       (NP *T*-3))))))))))
```

4.8.7 Illegal null elements

Following is a list of old and/or improperly formed null elements. They should be removed or updated as described, but they may occasionally slip into published files, despite checks which are designed to prevent this.

T should always have a reference index. Any *T* that lacks a reference index should be indexed as appropriate, or removed.

pseudo-attach used to be an all-purpose pseudo-attach marker. This should be converted to *ICH*, *RNR*, *EXP*, or *PPA*, as appropriate, or removed if now unnecessary. Note that the reference index for this null element appeared on the bracket label, making it appear to be an identity index.

T formerly stood for noun phrase *wh*-traces. It should be changed to *T* and indexed as appropriate. Note, however, that it is possible for 'T' to appear as part of real text (for example, as a symbol for 'Temperature' in scientific writing).

+ is inserted by the FIDDITCH parser as a passive trace. These are fixed in preprocessing, but those that slip through should be removed or replaced with '*'.

OF is inserted by the FIDDITCH parser after predeterminers such as *all*. It should be removed, along with its accompanying PP, and the noun phrases should be flattened into a single NP:

```
(NP (NP all)
    (PP OF
        (NP the dogs)))
```

should be changed to:

```
(NP all the dogs)
```

Note also that it is possible for a null element to still have a reference index after the brackets with the corresponding identity index have been removed. However, such errors are relatively rare (about 0.1% of all indexed null elements).

4.8.8 Limits of coindexation

The * null element generally receives a reference index whenever there is an appropriate referent elsewhere in the same sentence. However, there are cases in which annotators tend to not coindex, even when they can find a plausible referent. Some of these criteria overlap with each other and with rules described above.

Non-arguments. Annotators usually avoid indexing from non-arguments.

```
(S (PP for
       (NP us))
   ,
   (S-NOM-SBJ (NP-SBJ *)
```

```
           (VP eating
               (NP chocolate)))
   (VP is
       (NP-PRD (NP a way)
               (PP of
                   (NP life)))))


(S (PP For
       (NP Willie))
   ,
   (NP-SBJ (NP it)
           (S *EXP*-1))
   (VP is
       (ADJP-PRD difficult)
       (S-1 (NP-SBJ *)
            (VP to
                (VP resist
                    (NP chocolate))))))


(S (PP for (NP Zaphod))
   ,
   (NP-SBJ that steak)
   (VP is
       (ADJP-PRD ready
                 (SBAR (WHNP-1 0)
                       (S (NP-SBJ *)
                          (VP to
                              (VP eat
                                  (NP *T*-1))))))))
```

Gerund PP objects. Null subjects of gerund complements of PP modifiers of NPs are coindexed only if there is a particularly strong coindexed interpretation or the PP appears to be part of some "fixed phrase".

```
(S (NP-SBJ the company)
   (VP has
       (NP (NP no intention)
           (PP of
               (S-NOM (NP-SBJ *)
                      (VP tapping
                          (NP its short-term bank lines)))))
       (PP-TMP for
               (NP (NP a good part)
                   (PP of
                       (NP 1990))))))


(S (PP In
       (NP (NP addition)
           (PP to
               (S-NOM (NP-SBJ *-1)
                      (VP having
                          (NP high price-earnings ratios))))))
```

```
    ,
    (NP-SBJ-1 most)
    (VP pay
        (NP puny dividends)))
```

Possessive NPs. NP brackets that only mark a possessive phrase within an NP should NOT serve as a referent for a * null element:

```
(S (NP-SBJ (NP the Fed 's)
           goal)
   (VP is
       (S-PRD (NP-SBJ *)
              (VP to
                  (VP reduce
                      (NP inflation))))))
```

```
(NP (NP (NP Mr. Bush 's)
        claim
        (SBAR that...))
    and
    (NP (NP his insistence)
        (PP on
            (S-NOM (NP-SBJ *)
                   (VP combining...)))))
```

However, a possessive NP that is acting as a subject may serve such a role:

```
(S (NP-SBJ-1 Pinkerton 's)
   (VP had
       (VP locked
           (NP itself)
           (PP-CLR into
                   (NP low-price contracts))
           (S-PRP (NP-SBJ *-1)
                  (VP to
                      (VP win
                          (NP new business)))))))
```

Indexing in modified NPs. When an NP is adjoined with modifiers, the head NP should not serve as a referent for a * null element (although it may be used for template gapping). The adjunction level should also not serve as a reference for *'s found within the NP's modifiers.

In the following example, the * subject of *spotting* should not be indexed either to *New York money manager Mario Gabelli* or *an expert*, nor should it be indexed to the whole NP-SBJ.

```
(NP-SBJ (NP New York money manager Mario Gabelli)
        ,
        (NP (NP an expert)
            (PP at
                (S-NOM (NP-SBJ *)
                       (VP spotting
```

```
                                   (NP takeover candidates)))))
          ,)
```

In the following example, the * subject of *buy* should not be indexed, although it is clear that the *Soviet companies* are doing the buying.

```
(NP-SBJ-2 (NP Soviet companies)
          (VP needing
              (NP Western currencies)
              (S-PRP-CLR (NP-SBJ *)
                         (VP to
                             (VP buy
                                 (NP equipment and supplies)
                                 (ADVP-LOC abroad))))))
```

The following is an error, since the coindexation should have been from the whole subject noun phrase:

```
(S (NP-SBJ (NP-1 Ford)
           ,
           (SBAR (WHNP-64 which)
                 (S (NP-SBJ *T*-64)
                    (ADVP-TMP already)
                    (VP has
                        (NP an unwelcome (ADJP 13.2 %) holding))))
           ,)
   (VP is
       (ADJP-PRD prepared
                 (S (NP-SBJ *-1)
                    (VP to
                        (VP bid
                            (PP-PRP for
                                    (NP the entire company))))))))
```

5 Pseudo-Attach

5.1 Types of pseudo-attach

The pseudo-attach function is used for (1) structural ambiguity, (2) attachment in more than one place simultaneously, as with shared constituents, (3) indicating that something should be attached elsewhere, as with discontinuous dependencies, and (4) extraposed clauses. Each type of pseudo-attach is associated with a different type of null element (these are discussed in more detail in following sections; see also section 4 [Null Elements] for more information on indexing conventions):

1. Structural ambiguity *PPA* ("Permanent Predictable Ambiguity")

   Example: *I saw the man with the telescope*, where *PPA*-attach indicates an either/or interpretation at the attachment sites.

```
(S (NP-SBJ I)
   (VP saw
       (NP (NP the man)
           (PP *PPA*-1))
       (PP-MNR-1 with
                 (NP the telescope))))
```

2. Shared constituents *RNR* ("Right Node Raising")

   Example: *His dreams had revolved around her so much and for so long that...*, where *RNR*-attach indicates a simultaneous interpretation at the attachment sites.

```
(S (NP-SBJ His dreams)
   (VP had
       (VP revolved
           (PP-CLR around
                   (NP her))
           (UCP-ADV (ADVP (ADVP so much)
                          (SBAR *RNR*-1))
                     and
                     (PP-TMP for
                             (NP (NP so long)
                                 (SBAR *RNR*-1)))
                     (SBAR-1 that...)))))
```

3. Discontinuous dependency *ICH* ("Interpret Constituent Here")

   Example: *I saw a bear yesterday who was wearing really cool shoes*, where *ICH*-attach indicates that the relative clause is interpreted at the pseudo-attach site only.

```
(S (NP-SBJ I)
   (VP saw
       (NP (NP a bear)
           (SBAR *ICH*-2))
       (NP-TMP yesterday)
       (SBAR-2 (WHNP-1 who)
               (S (NP-SBJ *T*-1)
                  (VP was
```

```
                    (VP wearing
                        (NP (ADJP really cool) shoes)))))))
```

4. *it*-extraposition *EXP* ("EXPletive")

   Example: *My teacher said it was OK for me to use the notes on the test*, where *EXP*-attach indicates that the infinitive clause is the logical subject of the sentence.

```
(S (NP-SBJ My teacher)
   (VP said
       (SBAR 0
             (S (NP-SBJ (NP it)
                        (SBAR *EXP*-1))
                (VP was
                    (ADJP-PRD OK)
                    (SBAR-1 for
                            (S (NP-SBJ me)
                               (VP to
                                   (VP use
                                       (NP the notes)
                                       (PP-LOC on
                                               (NP the test))))))))))))
```

5.2 *PPA* ("Permanent Predictable Ambiguity")

This form of pseudo-attach is reserved for those cases in which one cannot tell even from context where a constituent should be attached. The **default** is to attach the constituent at the more likely site (or if that is impossible to determine, at the higher site) and then to pseudo-attach it at all other plausible sites.

Here, *on the printer* could modify either *the forms*, *the class or the forms*, or it could go directly under VP as a PP adverbial. The PP in question is adjoined to the NP *the forms* and *PPA*-attached to the other interpretation sites.

```
(S (NP-SBJ *)
   (VP Use
       (NP this option)
       (SBAR-TMP (WHADVP-2 when)
                 (S (NP-SBJ the operator)
                    (VP changes
                        (NP (NP (NP the class)
                                or
                                (NP (NP the forms)
                                    (PP-LOC-1 on
                                              (NP the printer))))
                            (PP-LOC *PPA*-1))
                        (PP-LOC *PPA*-1)
                        (ADVP-TMP *T*-2))))))
```

Here, *for the maintenance of COBOL* may modify *procedures* or it may function as a purpose clause attached under VP:

```
(S (NP-SBJ-2 (NP Procedures)
             (PP *PPA*-1))
   (VP have
       (VP been
           (VP established
               (NP *-2)
               (PP-PRP-1 for
                         (NP (NP the maintenance)
                             (PP of
                                 (NP COBOL))))))))
```

But since finding potential ambiguities is difficult and time-consuming, especially when reading in context, annotators are much more likely to attach at the most likely interpretation site than to use *PPA*-attach. Thus, use of *PPA* is very rare (about 40 occurrences in the WSJ corpus). (The alert reader will have noticed that *PPA* is as closely related to "Permanent Predictable Ambiguity" as it is to "Philadelphia Parking Authority").

5.2.1 Benign ambiguity

This use of *PPA* does not include cases of "benign ambiguity", and we do not in general show all plausible attachment points. In the case where a difference in attachment site makes no difference to the interpretation, the DEFAULT is to attach the constituent at the highest of the levels where it can be interpreted. Here, *in each record format* could modify the NP *different positions* as an adjoined postmodifier or it could modify the verb *begin* as an adverbial. The second of these interpretations is chosen since it represents the higher attachment site:

```
(S (NP-SBJ The key fields)
   (VP can
       (VP begin
           (PP-LOC-CLR at
                       (NP different positions))
           (PP-LOC in
                   (NP each record format)))))
```

5.2.2 Complex structural ambiguity

The *PPA* function can be used to indicate only simple attachment differences. The following are examples of ambiguity that involve more complex structural and categorical ambiguities which cannot be captured by *PPA*-attach. In this case, annotation proceeds according to the most likely interpretation.

- *I'll come by Tuesday*

```
(S (NP-SBJ I)
   (VP 'll
       (VP come
           (PRT by)
           (NP-TMP Tuesday))))
```

vs.

```
(S (NP-SBJ I)
   (VP 'll
       (VP come
           (PP-TMP by
                   (NP Tuesday)))))
```

- *I can't be happy long without drinking water*

```
(S (NP-SBJ-1 I)
   (VP ca n't
       (VP be
           (ADJP-PRD happy)
           (ADVP-TMP long)
           (PP without
               (S-NOM (NP-SBJ *-1)
                      (VP drinking
                          (NP water)))))))
```

vs.

```
(S (NP-SBJ I)
   (VP ca n't
       (VP be
           (ADJP-PRD happy)
           (ADVP-TMP long)
           (PP without
               (NP drinking water)))))
```

5.2.3 Further examples of *PPA*

```
(S (PP-LOC In
           (NP Southern Illinois))
   ,
   (NP-SBJ-2 (NP the new federal program)
             (PP of
                 (NP (NP help)
                     (PP to
                         (NP (ADJP economically depressed)
                             areas)))))
   (VP ought
       (S (NP-SBJ *-2)
          (VP to
              (VP provide
                  (NP (NP some stimulus)
                      (PP *PPA*-1))
                  (PP-CLR-1 to
                            (NP growth)))))))
```

```
(S (NP-SBJ *)
```

```
   (VP Imagine
       (NP (NP (NP the searching)
               and
               (NP the prayer))
           (SBAR (WHNP-1 that)
                 (S (NP-SBJ *T*-1)
                    (VP lay
                        (PP-LOC-CLR behind
                                    (NP (NP the letter)
                                        (SBAR (WHNP-2 0)
                                              (S (NP-SBJ the rector)
                                                 (VP wrote
                                                     (NP *T*-2)
                                                     (PP-TMP *PPA*-3))))))
                        (PP-TMP-3 after
                                  (NP (NP (QP almost a) decade)
                                      (PP of
                                          (NP (NP service)
                                              (PP to
                                                  (NP this majestic
                                                      church))))))))))))

(S (NP-SBJ-2 There)
   (VP ought
       (S (NP-SBJ *-2)
          (VP to
              (VP be
                  (NP-PRD (NP a point)
                          (SBAR (WHPP-1 beyond
                                        (WHNP which))
                                (S (NP-SBJ we)
                                   (VP will not
                                       (VP allow
                                           (S (NP-SBJ ourselves)
                                              (VP to
                                                  (VP go
                                                      (PP-DIR *T*-1))))
                                           (ADVP *PPA*-3)))))))))
       (ADVP-3 regardless
               (PP of
                   (SBAR-NOM (WHNP-4 what)
                             (S (NP-SBJ Russia)
                                (VP does
                                    (NP *T*-4))))))))

(S (NP-SBJ-2 You)
   (VP must
       (VP specify
           (NP the link type)
           (PP-TMP before
                   (S-NOM (NP-SBJ *-2)
                          (VP using
```

```
                             (NP the START command)
                             (S-CLR *PPA*-1))))
            (S-PRP-1 (NP-SBJ *)
                     (VP to
                         (VP activate
                             (NP the link)))))))
```

5.3 *RNR* ("Right Node Raising")

This type of pseudo-attach is used for those cases in which a constituent should be interpreted simultaneously in more than one place. (See section 8 [Shared Complements and Modifiers] for more on shared constituents.)

```
(S But
   (NP-SBJ-2 our outlook)
   (VP (VP has
           (VP been
               (ADJP-PRD *RNR*-1)))
       ,
       and
       (VP continues
           (S (NP-SBJ *-2)
              (VP to
                  (VP be
                      (ADJP-PRD *RNR*-1)))))
       ,
       (ADJP-PRD-1 defensive)))


(S (NP-SBJ One)
   (VP knows
       (ADVP-CLR better)
       ,
       (ADVP-TMP now)
       ,
       (SBAR (SBAR (WHNP-2 who)
                   (S (NP-SBJ *T*-2)
                      (VP has
                          (NP (NP bone)
                              (PP-LOC *RNR*-4)))))
             and
             (SBAR (WHNP-3 who)
                   (S (NP-SBJ *T*-3)
                      (VP has
                          (NP (NP jelly)
                              (PP-LOC *RNR*-4)))))
             (PP-LOC-4 in
                       (NP his spine)))))


(NP (NP (ADJP so many) enchained demons)
    (VP straining
        (PP-MNR in
                (NP anger))
```

```
(S (NP-SBJ *)
   (VP to
       (VP (VP tear
               (NP *RNR*-1))
           and
           (VP gnaw
               (PP-CLR on
                       (NP *RNR*-1)))
           (NP-1 his bones))))))
```

5.4 *ICH* ("Interpret Constituent Here")

The most common type of pseudo-attach is *ICH*-attach, which is used to indicate a relationship of constituency between elements separated by intervening material. For instance, *ICH*-attach is used in "heavy shift" constructions when the movement results in a configuration in which it is impossible to attach the constituent to the phrase it belongs with:

```
(S (NP-SBJ (NP a young woman)
           (SBAR *ICH*-1))
   (VP entered
       (SBAR-1 (WHNP-2 whom)
               (S (NP-SBJ she)
                  (PP-TMP at
                          (ADVP once))
                  (VP recognized
                      (NP *T*-2)
                      (PP-CLR as
                              (NP Jemima Broadwood)))))))
```

5.4.1 Word order

ICH-attach is never used solely to indicate **word order**; there must also be a difference in *attachment height*. For example, the following example does not require *ICH*-attach of the NP containing *a very nice mermaid* (here, because the sentence adverbial is attached in VP):

```
(S (NP-SBJ I)
   (VP met
       (PP-LOC at
               (NP the dock))
       (NP (NP a
               (ADJP very nice)
               mermaid)
           (SBAR (WHNP-2 who)
                 (S (NP-SBJ-3 *T*-2)
                    (VP offered
                        (S (NP-SBJ *-3)
                           (VP to
                               (VP take
                                   (NP me)
                                   (PP-CLR for
                                           (NP a swim))))))))))
```

However, if the wording were somewhat different, *ICH*-attach would be appropriate. Here, the adverbial intervenes between a head and its relative clause, making *ICH*-attach necessary:

```
(S (NP-SBJ I)
   (VP met
       (NP (NP a
               (ADJP very nice)
               mermaid)
           (SBAR *ICH*-1))
       (PP-LOC at
               (NP the dock))
       (SBAR-1 (WHNP-2 who)
               (S (NP-SBJ-3 *T*-2)
                  (VP offered
                      (S (NP-SBJ *-3)
                         (VP to
                             (VP take
                                 (NP me)
                                 (PP-CLR for
                                         (NP a swim))))))))))
```

Likewise, *ICH*-attach is not used to show deviations from canonical word order in constructions in which the constituency is not affected and where the appropriate attachment levels are available:

```
(S (NP-SBJ I)
   (VP put
       (PP-PUT on
               (NP the shelf))
       (NP (NP that book)
           (SBAR (WHNP-2 that)
                 (S (NP-SBJ you)
                    (VP loaned
                        (NP me)
                        (NP *T*-2)
                        (NP-TMP the other day)))))))
```

5.4.2 Further examples

```
(S (NP-SBJ-2 (NP Nothing)
             (PP *ICH*-1))
   (VP was
       (S (NP-SBJ-3 *-2)
          (VP to
              (VP be
                  (VP seen
                      (NP *-3)))))
       (PP-1 but
             (NP water))))
```

```
(S (NP-SBJ-3 Another attempt
```

```
          (S *ICH*-1))
   (VP will
       (VP be
           (VP made
               (NP *-3)
               (NP-TMP this year)
               (PP-LOC in
                       (NP New Orleans))
               (S-2 (NP-SBJ *)
                    (VP to
                        (VP resume
                            (NP the program))))))))

(S (NP-SBJ (NP A stranger)
           (NP *ICH*-1))
   (VP was
       (PP-LOC-PRD before
                   (NP him))
       --
       (NP-1 (NP a boy)
             (ADJP (ADJP (NP a shade)
                          larger)
                   (PP than
                       (NP himself))))))

(S I could not help calling to mind my little brother's face,
   (SBAR-TMP (WHADVP-2 when)
             (S (NP-SBJ (NP he)
                        (NP *ICH*-1))
                (VP was
                    (VP sleeping
                        (ADVP-TMP *T*-2)
                        (NP-1 (NP an infant)
                              (PP-LOC in
                                      (NP the cradle))))))))

(S (NP-SBJ he)
   (VP wondered
       (SBAR (WHADVP-2 where)
             (S (NP-SBJ the superstition
                        (SBAR *ICH*-1))
                (VP had
                    (VP originated
                        (ADVP-LOC *T*-2)
                        (SBAR-1 that
                                (S it was bad luck for... ))))))))

(S (ADVP However)
   ,
   (NP-SBJ it)
   (VP does
```

```
(VP require
    (NP (NP more personal computer memory)
        (PP *ICH*-1))
    (S-CLR (NP-SBJ *)
           (VP to
               (VP run
                   (NP *?*))))
    (PP-1 than
          (NP type 0)))))
```

5.4.3 Conjunctive Prepositional Phrases

In cases such as the following, the PP should not be pseudo-attached. Unfortunately, this policy is a bit counter-intuitive to the annotators, so they are occasionally bracketed with *ICH*-attach anyway. (For example, more than a third of such *"including"* constructions in the WSJ corpus were bracketed with *ICH*.)

- *with*

```
(S (NP-SBJ (NP the letters)
           (SBAR (WHNP-5 0)
                 (S (NP-SBJ you)
                    (VP specify
                        (NP *T*-5)))))
   (VP follow
       (PP with
           (NP their definitions))))
```

- *including*

```
(S (NP-SBJ Several ridiculous projects)
   (VP continue
       ,
       (PP including
           (NP the New International Economic Order))))


(S (NP-SBJ (NP Several ridiculous projects)
           (PP *ICH*-1))
   (VP continue
       ,
       (PP-1 including
             (NP the New International Economic Order))))
```

5.5 *EXP* ("EXPletive")

In cases where a clausal subject has been extraposed and replaced by an expletive *it*, we use a type of pseudo-attach called *EXP*. (In the small ATIS sample included with this release, it is also used for existential *there*.) Use of *EXP*-attach is discussed in more detail in section 17 [*It*-Extraposition].

```
(S (NP-SBJ (NP It)
           (SBAR *EXP*-1))
   (VP is
       (ADJP-PRD clear)
       (PP to
           (NP me))
       (SBAR-1 that
               (S (NP-SBJ this message)
                  (VP is
                      (ADJP-PRD unclear))))))
```

```
(S (PP To
       (NP Flavia))
   (NP-SBJ (NP it)
           (S *EXP*-1))
   (VP is
       (ADJP (ADJP more necessary)
             (SBAR *ICH*-3))
       (S-1 (NP-SBJ-2 *)
            (VP to
                (VP be
                    (VP called
                        (S (NP-SBJ *-2)
                           (ADJP-PRD clever))))))
       (SBAR-3 than
               (S (NP-SBJ *)
                  (VP to
                      (VP breathe))))))
```

```
(S (NP-SBJ (NP It)
           (S *EXP*-1))
   (VP is
       (ADJP-PRD easy)
       (S-1 (NP-SBJ *)
            (VP to
                (VP see
                    (SBAR (WHADVP-2 why)
                          (S (NP-SBJ the ancient art)
                             (VP is
                                 (PP-LOC-PRD on
                                             (NP the ropes))
                                 (ADVP-PRP *T*-2)))))))))
```

5.6 Punctuation

Punctuation should never be pseudo-attached.

"Two techniques are available to accomplish the platform heading: the use of external or surveying equipment to establish the proper heading; the use of the character of the platform components for an indication of true heading."

```
(S (NP-SBJ (NP two techniques)
           (NP *ICH*-4))
   (VP are
       (ADJP-PRD available...)
       :
       (NP-4 (NP the use of...)
             ;
             (NP the use of...))))
```

6 Copular Verbs

6.1 Simple copular complements

Complements of the following verbs appear with the -PRD tag. This list should be considered exhaustive (see [Quirk et al. 1985] sections 16.21-24).

be (friendly/my friend/at home) [adj/n/adv]

appear (happy/the only solution) [adj/n]
feel (annoyed/a fool) [adj/n]
look (pretty/a fine day) [adj/n]
seem (restless/a genius) [adj/n]
smell (sweet) [adj]
sound (surprised/a reasonable idea) [adj/n]
taste (bitter) [adj]

remain (uncertain/good friends) [adj/n]
keep (silent) [adj]
stay (motionless/good friends) [adj/n]

become (older/an expert) [adj/n]
come (true) [adj]
end up (happy/her slave) [adj/n]
get (ready) [adj]
go (sour) [adj]
grow (tired) [adj]
prove (rather useful/his equal) [adj/n]
turn (cold/traitor) [adj/n]
turn out (fortunate/a success) [adj/n]
wind up (drunk/a millionaire) [adj/n]

6.1.1 Adjectival

```
(S (NP-SBJ The dog)
   (VP is/appears/seems
       (ADJP-PRD happy)))
```

```
(S (NP-SBJ That food)
   (VP looks/smells/tastes
       (ADJP-PRD awful)))
```

```
(S (INTJ Please)
   (NP-SBJ *)
   (VP keep/remain/stay
       (ADJP-PRD silent)))
```

This includes "pseudo-adjectives" (see section 15 [Small Clauses]).

```
(S (NP-SBJ Things)
   (VP seem
       (PP-PRD under
               (NP control))))
```

6.1.2 Nominal

```
(S (NP-SBJ Your safety belt)
   (VP is-PRD your friend))


(S (NP-SBJ-1 The former chief executive)
   (VP will
       (VP remain
           (NP-PRD chairman)))
   .)


(S (NP-SBJ The new student)
   (VP proved
       (NP-PRD an idiot)))
```

but note:

```
(S (NP-SBJ The new student)
   (VP proved
       (NP a theorem)))
```

6.1.3 Adverbial

Adverbial predicates should only be tagged -PRD when they follow *be* or in a *do so* construction.

- after *be*

```
(S (NP-SBJ (NP much)
           (PP of
               (NP the action)))
   (VP was
       (PP-LOC-PRD in
                   (NP heating oil))))


(S (NP-SBJ The party)
   (VP will
       (VP be
           (PP-TMP-PRD at
                       (NP eleven)))))


(S (NP-SBJ business)
   (VP is
       (ADVP-PRD up
                 (NP 35 %))
       (PP-TMP in
               (NP the past year))))
```

- *do so* constructions

```
(S (SBAR-ADV while the SEC regulates who files)
   ,
   (NP-SBJ the law)
   (VP tells
       (NP them)
       (SBAR (WHADVP-1 when)
             (S (NP-SBJ *)
                (VP to
                    (VP do
                        (ADVP-PRD so)
                        (ADVP-TMP *T*-1)))))))


(S (S (NP-SBJ I)
      (VP attend))
   , and
   (SINV (ADVP-PRD-TPC-1 so)
         (VP does
             (ADVP-PRD *T*-1))
         (NP-SBJ (NP a television crew)
                 (PP from
                     (NP New York City)))))
```

However, adverbial modifiers are sometimes erroneously tagged -PRD in sentences such as *Mandela remains in prison*.

6.1.4 Clausal

Forms of *be* can also take clausal complements:

```
(S (NP-SBJ its purpose)
   (VP is
       (S-PRD (NP-SBJ *)
              (VP to
                  (VP gauge
                      (NP learning progress))))))


(S (NP-SBJ The theory)
   (VP was
       (SBAR-PRD that
                 (S (NP-SBJ the Voice)
                    (VP is
                        (NP-PRD a propaganda agency))))))
```

including some fairly unusual ones:

```
(S (ADVP-TMP Now)
   (NP-SBJ the question)
   (VP is :
       (SQ-PRD Is
               (NP-SBJ Poland)
```

```
                (ADJP-PRD ready
                          (PP for
                              (NP it))))))
```

However, when *be* acts as a semimodal, the following S should not get -PRD:

```
(S (NP-SBJ-1 You)
   (VP are
       (S (NP-SBJ *-1)
          (VP to
              (VP resign
                  (ADVP-TMP immediately))))))
```

6.2 Related constructions

It is noted in [Quirk et al. 1985] that, for many of the above verbs with nominal complements, English speakers (especially Americans) tend to (strongly) prefer a variant containing *to be* or *like* instead of just a simple NP. The Treebank treats such constructions as follows:

6.2.1 The "like" versions

Verbs that take a complement mediated by *like* should be bracketed with a -CLR tag rather than -PRD.

```
(S (NP-SBJ That)
   (VP sounds
       (PP-CLR like
               (NP a reasonable idea))))


(S (NP-SBJ It)
   (VP looks
       (PP-CLR like
               (NP a fine day))))
```

6.2.2 The "to be" versions

Versions with *to be* are bracketed as complement clauses (as described in section 15 [Small Clauses], under "Null subject clausal complements").

```
(S (NP-SBJ-3 That)
   (VP appears
       (S (NP-SBJ *-3)
          (VP to
              (VP be
                  (NP-PRD the only solution))))))
```

7 Coordination

7.1 Basic information on the level of coordination

Coordination in all cases is done at the lowest possible level.

7.1.1 Word-level

Single-word elements of the same syntactic category are coordinated at word-level and are annotated with flat structure.

```
(NP John and Mary)


(NP Susan , John and Mary)


(S (NP-SBJ Baking and eating)
   (VP are
       (ADJP-PRD fun)))


(S (S-NOM-SBJ (NP-SBJ *)
              (VP Baking
                  and
                  eating
                  (NP cookies)))
   (VP is
       (ADJP-PRD fun)))
```

7.1.2 Phrase-level

1. When one or more of the coordinated elements are multi-word (and of the same syntactic type), each element is bracketed with the appropriate label, as is the immediately dominating node. All coordinating conjunctions are children of the top phrase node ("conjunction level" or "coordination level"). (See section 8 [Shared Complements and Modifiers] for more information.)

```
(NP (NP three cookies)
    and
    (NP a book))


(NP (NP cookies)
    and
    (NP a book))


(VP (VP fishing)
    and
    (VP tying
        (NP flies)))
```

2. Exception to the above: coordination of premodifiers in within NP.

   The annotation of noun phrases is quite complex, however, and follows somewhat different guidelines than for the other categories. See section 8 [Shared Complements and Modifiers] for information on coordination in noun phrases.

7.2 Coordination of unlike syntactic categories (UCP)

In general, coordinated phrases belong to the same syntactic category. However, it is also possible for coordinated phrases to belong to different categories. When they do, the phrase node at the level of coordination is labeled UCP ("Unlike Coordinated Phrase").

7.2.1 Outside of NPs

Unlike the coordination of single words of the same syntactic category, the coordination of single words of different syntactic categories is often shown with phrase-level coordination. When the coordination is not of modifiers in a noun phrase (see section 7.2.2 below for UCP in NPs), each conjunct is given its own appropriate bracket label, and the outer coordination is labeled UCP.

```
(S (NP-SBJ U.S. interest)
   (VP may
       (VP be
           (UCP-PRD (ADJP-PRD big)
                    and
                    (VP growing)))))

(S-NOM (NP-SBJ *)
       (VP serving
           (NP (NP the wishes)
               (PP of (NP the client)))
           (UCP-MNR (ADVP fairly)
                    and
                    (PP in
                        (NP an efficient manner)))))
```

UCP may support function tags as well, but only in the case where the function tag applies to all conjuncts.

- ```
  (S-NOM (NP-SBJ *)
         (VP serving
             (NP (NP the wishes)
                 (PP of (NP the client)))
             (UCP-MNR (ADVP fairly)
                      and
                      (PP in
                          (NP an efficient manner)))))
  ```

- ```
  (S And
     (ADVP-TMP now)
     ,
     (NP-SBJ-1 the woman)
     ,
  ```

```
    (S-ADV (NP-SBJ *-1)
           (UCP-PRD (ADJP-PRD tired)
                    and
                    (VP trembling)))
    ,
    (VP came
        (ADVP-DIR here)
        (PP-DIR to
                (NP the DeKalb County cannery)))
    .)
```

- ```
  (S (NP-SBJ America West)
     ,
     (ADVP though)
     ,
     (VP is
         (UCP-PRD (NP a smaller airline)
                  and
                  (ADVP therefore)
                  (ADJP (ADJP more affected)
                        (PP by
                            (NP (NP the delayed delivery)
                                (PP of
                                    (NP a single plane))))
                        (SBAR than
                              (S (NP-SBJ (NP many)
                                         (PP of
                                             (NP its competitors)))
                                 (VP would
                                     (VP be
                                         (ADJP *?*))))))))
     .)
  ```

- ```
  (S (S-ADV (NP-SBJ *-1)
            (UCP-PRD (ADJP (NP Seven years)
                           late
                           (PP in
                               (NP the launching)))
                     ,
                     (PP (NP (QP $ 1 billion) *U*)
                         over
                         (NP budget))
                     and
                     (NP (NP a target)
                         (PP of
                             (NP anti-nuclear protestors))))
            ,)
     (NP-SBJ-1 Galileo)
     (VP has
         (VP (ADVP-TMP long)
             been
             (NP-PRD (NP a symbol)
  ```

```
                    (PP of
                        (NP trouble)))))
          .)
```

- ```
  (S (NP-SBJ His plans and dreams)
     (VP had
         (VP revolved
             (PP-LOC-CLR around
                         (NP her))
             (UCP (ADVP (ADVP so much)
                        (SBAR *RNR*-1))
                  and
                  (PP-TMP for
                          (ADVP (ADVP so long)
                                (SBAR *RNR*-1)))
                  (SBAR-1 that
                          (S (ADVP-TMP now)
                             (NP-SBJ he)
                             (VP felt
                                 (SBAR as if
                                       (S (NP-SBJ he)
                                          (VP had
                                              (NP nothing))))))))))
  ```

7.2.2 In NPs

As usual, NP structure is different from the structure of other categories. When elements that belong to different syntactic categories are coordinated, and that coordinated structure modifies an NP, the structure is labeled UCP.

If both conjuncts are single words, no structure is shown inside the UCP. Note that this is generally not the case outside of NPs — in other categories, single-word conjuncts are labeled inside UCP.

```
(NP (UCP federal and state) laws)
```

Multi-word adjectival modifiers within the UCP should be labeled. Multi-word nominal modifier within the UCP are often unlabeled. (See section 8 [Shared Complements and Modifiers] and section 11 [Modification of NP] for more information on nominal modifiers.)

```
(NP the
    (UCP highest quality
         and
         (ADJP most reasonably priced))
    product)
```

7.3 General guidelines for the bracketing of coordinated structures

7.3.1 Labeling at the level of coordination

A conjunction (sometimes unlabeled, sometimes CONJP) joins two or more elements, including those that are typologically different and thus dominated by UCP ("Unlike Coordinated Phrase"). Phrases with identical bracket labels or part-of-speech tags are of course coordinated under the appropriate bracket label (e.g., the level of coordination for NPs and for single-word nouns is labeled NP, etc.).
```

7.3.2 Function tags at the level of coordination

Function tags appear only on the bracket label at the highest level of coordination in coordinated phrases of the same phrase type and the same function.

```
(S The Burger King worker handled the order
   (ADVP-MNR (ADVP very slowly)
             but
             (ADVP not very carefully)))


(S-NOM (NP-SBJ *)
       (VP serving
           (NP (NP the wishes)
               (PP of (NP the client)))
           (UCP-MNR (ADVP fairly)
                    and
                    (PP in
                        (NP an efficient manner)))))
```

However, if all conjuncts do not share the same function, the function tags appear on individual conjuncts.

```
(S (PP (PP-TMP After
               (NP the 1987 crash))
       , and
       (PP-PRP as
               (NP (NP a result)
                   (PP of
                       (NP (NP the recommendations)
                           (PP of
                               (NP many studies)))))))

   ``circuit breakers'' were devised to allow market participants to
   regroup and restore orderly market conditions)


(S But in the ``Bare-Faced Messiah'' case the author found most of his
   material
   (PP (PP-LOC in
               (NP court records))
       or
       (PP-MNR via
               (NP the
                   (NAC Freedom
                        (PP of
                            (NP Information)))
                   Act))))
```

7.3.3 Internal structure

1. The internal structure of the coordinated phrase is left flat if possible. For example, the internal structure of the coordination of single-words is not shown and thus only the level of coordination is labeled.

```
(NP apples and oranges)
```

```
(NP the apples and oranges)
```

```
(S (NP-SBJ Curious George)
   (VP cut and pasted
       (NP the pictures)))
```

```
(S (NP-SBJ I)
   (VP slept and dreamed))
```

2. If at least one of the coordinated phrases is multi-word, coordination is at the lowest possible level (i.e., the lowest level that includes the shared constituent and excludes any unshared constituents).

```
(S (NP-SBJ I)
   (VP (VP slept)
       and
       (VP dreamed
           (SBAR that
                 (S (NP-SBJ I)
                    (VP was
                        (NP-PRD a butterfly)))))))
```

3. NX is used to show the internal structure of coordinate structures with multi-word conjuncts in NPs. There are no corresponding "X" levels for other parts of speech. See section 8 [Shared Complements and Modifiers] and for more on the use of NX.

```
(NP the
    (NX (NX red book)
        and
        (NX yellow pencils)))
```

4. An extended example.

   This is an example of the minimal levels used for the coordination of verb phrases.

   (a) V-level.

   If all conjoined verbs are single-word and all objects and/or modifiers are shared, coordination is at word level.

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Baking
                  and
                  eating
                  (NP cookies)))
   (VP is
       (ADJP-PRD fun)))
```

   (b) VP-level.

i. Unshared objects and modifiers.
If there are unshared objects or modifiers, coordination is at the lowest possible VP-level.

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP (VP Baking
                      (NP pies))
                  and
                  (VP eating
                      (NP cookies))))
   (VP are
       (NP-PRD fun activities)))

(S (S-NOM-SBJ (NP-SBJ *)
              (VP (VP Baking
                      (PP-LOC in
                              (NP (NP Grandma's)
                                  kitchen)))
                  and
                  (VP eating
                      (PP-LOC in
                              (NP bed)))))
   (VP are
       (ADJP-PRD fun)))

(S (NP-SBJ John)
   (VP will
       (VP (VP have
               (VP baked
                   (NP a cake)))
           and
           (VP have
               (VP frosted
                   (NP the cupcakes)))
           (NP-TMP this morning))))
```

ii. Shared objects and modifiers with multi-word conjuncts.
Coordination is at the level of the lowest possible VP, and *RNR*-attach is used when a shared object must be attached at different levels of structure. Shared modifiers are not *RNR*-attached. (See section 8 [Shared Complements and Modifiers] for more on *RNR*.)

```
(S (NP-SBJ-1 *)
   (VP Do
       (VP avoid
           (S (NP-SBJ-2 *-1)
              (VP (VP puncturing
                      (NP *RNR*-5))
                  or
                  (VP cutting
                      (PP-CLR into
                              (NP *RNR*-5)))
                  (NP-5 meats)
                  (S-PRP (NP-SBJ *-2)
                         (VP to
                             (VP test
                                 (NP them)))))))))
```

```
(S (NP-SBJ John)
   (VP will
       (VP (VP have
               (VP baked
                   (NP *RNR*-1)))
           and
           (VP have
               (VP frosted
                   (NP *RNR*-1)))
           (NP-1 the cake))))
```

(c) S-level.

Coordination is at S-level only in the case where each conjunct has an overt subject (whether coreferential or not).

```
(S (NP-SBJ She)
   (VP was
       (VP eating
           (ADVP-MNR quietly)
           (S-ADV (S (NP-SBJ her head)
                     (VP hanging))
                  ,
                  and
                  (S (NP-SBJ her scaly, dead-looking foot)
                     (VP lifted
                         (NP-ADV just a little)
                         (PP-CLR from
                                 (NP the ground))))))))
```

```
(S (NP-SBJ I)
   (VP admire
       (S (S (NP-SBJ your)
             (VP hanging
                 (NP your head)))
          and
          (S (NP-SBJ your)
             (VP dangling
                 (NP your feet))))))
```

7.4 Gapping

This section presents the official annotation policy for gapped structures, but also describes some likely common alternates found in the corpus.

7.4.1 VP gapping

In the case of VP gapping, when the second conjunct lacks both subject and verb, coordination is at VP-level and gap-coindexing (also referred to as template gapping) is used. (See section 1 [Overview of Basic Clause Structure] and section 4 [Null Elements] for more on the template approach to gapping.)

```
(S (NP-SBJ I)
   (VP (VP cooked
```

```
            (NP-6 soup)
            (NP-TMP-7 Wednesday))
        and
        (VP (NP=6 potatoes)
            (NP-TMP=7 Thursday)))
    .)


(S (NP-SBJ John)
   (VP (VP is
           (ADVP-TMP-5 sometimes)
           (NP-PRD-6 a gentleman))
       but
       (VP (ADVP-TMP=5 usually)
           (NP-PRD=6 a child))))


(S (NP-SBJ John)
   (VP (VP is
           (ADJP-PRD-7 endearing))
       and
       (VP (ADVP-TMP sometimes)
           (ADJP-PRD=7 amusing))))
```

A common alternate to VP gapping in these cases (i.e., when the second conjunct lacks both a verb and a subject), especially when a modifier occurs only in the second conjunct, is to coordinate the predicates, leaving the modifier at conjunction level. Therefore one might find two of the above sentences bracketed as follows:

```
(S (NP-SBJ John)
   (VP is
       (NP-PRD (ADVP-TMP sometimes)
               (NP a gentleman)
               but
               (ADVP-TMP usually)
               (NP a child))))


(S (NP-SBJ John)
   (VP is
       (ADJP-PRD (ADJP endearing)
                 and
                 (ADVP-TMP sometimes)
                 (ADJP amusing))))
```

The problem is that ADVPs inside ADJPs are generally regarded as simple intensifiers, such as *very, somewhat* or *a little,* and *sometimes* here is an "S-level" temporal modifier. The official solution is to go up a level of coordination and gap, but some annotators feel this is excessive, and would like a simpler way that still gets *sometimes* interpreted at a version of "S-level".

Below, *in general* and *in particular* are interpreted as adjuncts to NP, rather than sentential; consequently template gapping is **not** used.

```
(S (NP-SBJ *-1)
   (VP spurred
       (PRT on)
       (PP by
```

```
            (NP-LGS (NP (NP the growth)
                        (PP of
                            (NP government))
                        (PP in
                            (ADJP general)))
                    and
                    (NP (NP (NP budget gimmicks)
                            and
                            (NP deceptive management))
                        (PP in
                            (NP particular))))))
    .)
```

7.4.2 Gapping in a small clause

Gap-coindexing may be used in a small clause when there is a modifier with one or both of the predicates. As with VP gapping, the official policy is to gap as in the first example, but the second example represents a common alternative.

```
(SBAR (WHNP-8 who)
      (S (NP-SBJ we)
         (VP find
             (S (S (NP-SBJ *T*-8)
                   (ADJP-PRD-1 endearing))
                and
                (S (ADVP-TMP sometimes)
                   (ADJP-PRD=1 amusing))))))

(SBAR (WHNP-9 who)
      (S (NP-SBJ we)
         (VP find
             (S (NP-SBJ *T*-9)
                (ADJP-PRD (ADJP endearing)
                          and
                          (ADVP-TMP sometimes)
                          (ADJP amusing))))))
```

7.4.3 Gapping at S-level

The level of coordination in a gapping structure must be S-level when both a subject and a VP element (e.g., VP, object, or adjunct) are present in the second conjunct.

- Subject and VP

```
(NP fears
    (SBAR that
          (S (S (NP-SBJ-2 the Thatcher government)
                (VP may
                    (VP be
                        (PP-PRD-3 in
                                  (NP turmoil)))))
             and
```

```
(S (NP-SBJ=2-1 (NP Britain 's)
               Labor Party)
   (VP=3 positioned
         (S (NP-SBJ *-1)
            (VP to
                (VP regain
                    (NP (NP control)
                        (PP of
                            (NP the
                                government))))))))))
```

- Subject and object:

```
(S (S (NP-SBJ-10 Mary)
      (VP likes
          (NP-11 potatoes)))
   and
   (S (NP-SBJ=10 Bill)
      ,
      (NP=11 ostriches)))
```

- Subject and adjunct:

```
(S (S (NP-SBJ-1 Mary)
      (VP likes
          (S (NP-SBJ *-1)
             (VP to
                 (VP swim
                     (PP-TMP-2 on
                               (NP Tuesdays)))))))
   and
   (S (NP-SBJ=1 Bill)
      (PP-TMP=2 on
                (NP Wednesdays))))
```

7.4.4 Gapping in PP

Coordination of NPs (as in the first example) is a likely alternate to gapping within PPs (as in the second example).

```
( (S (NP-SBJ-1 The Nashville plan)
     (VP has
         (VP become
             (VP recognized
                 (NP *-1)
                 (PP-CLR as
                         (NP (NP (ADVP perhaps)
                                 (NP the
                                     (ADJP most acceptable))
                                 and
                                 (ADVP thus)
```

```
                                        (NP the
                                            (ADJP most practical)))
                                    (SBAR (WHNP-2 0)
                                          (S (NP-SBJ *)
                                             (VP to
                                                 (VP put
                                                     (NP *T*-2)
                                                     (PP-CLR into
                                                             (NP effect))
                                                     (PP-LOC in
                                                             (NP the
                                                                 troubled
                                                                 South))))))))))
     .))

( (S (NP-SBJ-1 The Nashville plan)
     (VP has
         (VP become
             (VP recognized
                 (NP *-1)
                 (PP-CLR (PP as
                             (ADVP-3 perhaps)
                             (NP-4 (NP the
                                       (ADJP most acceptable))
                                   (SBAR *RNR*-5)))
                         and
                         (PP (ADVP=3 thus)
                             (NP=4 (NP the
                                       (ADJP most practical))
                                   (SBAR *RNR*-5)))
                         (SBAR-5 (WHNP-2 0)
                                 (S (NP-SBJ *)
                                    (VP to
                                        (VP put
                                            (NP *T*-2)
                                            (PP-CLR into
                                                    (NP effect))
                                            (PP-LOC in
                                                    (NP the
                                                        troubled
                                                        South)))))))))))
     .))
```

7.4.5 Gapping in NP

```
(S (NP-SBJ-111 The remaining
               (QP $ 21.9 billion)
               *U*)
   (VP could
       (VP be
           (VP raised
```

```
            (NP *-111)
            (PP through
                (NP (NP (NP the sale)
                        (PP of
                            (NP-2 short-term Treasury bills))
                        (PP-3 *NOT*))
                    ,
                    (NP (NP (NP=2 two-year notes)
                            (PP-TMP=3 in
                                      (NP November)))
                        and
                        (NP (NP=2 five-year notes)
                            (PP-TMP=3 in
                                      (NP early December)))))))))
 .)
```

7.4.6 Some difficult cases

1. *NOT*

   One may occasionally find annotation such as the following, where the anti-placeholder *NOT* indicates that certain items found in the second conjunct are not interpreted in the first. See section 4 [Null Elements] for more on the use of *NOT*. For example, here *perhaps* and the *if*-clause go only with the second conjunct, indicated by *NOT* in the first conjunct.

```
(S (NP-SBJ (NP Robert Schneider)
           (PP of
               (NP Duff & Phelps)))
   (VP sees
       (S (NP-SBJ paper-company stock prices)
          (VP (VP falling
                  (NP-EXT-2 (NP 10 %)
                            to
                            (NP 15 %))
                  (PP-TMP in
                          (NP 1990))
                  (ADVP-1 *NOT*)
                  (SBAR-ADV-3 *NOT*))
              ,
              (VP (ADVP=1 perhaps)
                  (NP-EXT=2 25 %)
                  (SBAR-ADV=3 if
                              (S (NP-SBJ there)
                                 (VP 's
                                     (NP-PRD a recession)))))))))
```

2. Template gapping with *RNR*-attach.

   RNR-attach may be used in gapping structures for non-complements that are shared across clauses. In the following example, the adjunct (*in the rush-hour tremor...*) is extracted from a subordinate clause in the first conjunct but from a main clause in the second. Note that *RNR*-attach is not usually used with adjuncts but is used here because the clauses over which gapping takes place are not symmetrical. See section 8 [Shared Complements and Modifiers] for more on the use of *RNR*.

```
(S (S (NP-SBJ-3 (QP At least 270)
                people)
      (VP were
          (VP reported
              (S (NP-SBJ-1 *-3)
                 (VP-2 killed
                       (NP *-1)
                       (PP-LOC *RNR*-5))))))
   and
   (S (NP-SBJ=3-6 1,400)
      (VP=2 injured
            (NP *-6)
            (PP-LOC *RNR*-5)))
   (PP-LOC-5 in
             (NP (NP the rush-hour tremor)
                 (SBAR (WHNP-4 that)
                       (S (NP-SBJ *T*-4)
                          (VP caused
                              (NP (NP billions)
                                  (PP of
                                      (NP (NP dollars)
                                          (PP of
                                              (NP damage)))))
                              (PP-LOC along
                                      (NP (NP 100 miles)
                                          (PP of
                                              (NP the
                                                  San Andreas fault))))))))))
```

7.5 Bracketing of coordinating conjunctions

7.5.1 Coordinating conjunctions

1. Single-word.

   Single-word conjunctions are unlabeled.

   and, but, or
   yet, so
   less, plus, minus

```
(NP (NP a hammer)
    and
    (NP a nail))
```

2. Multi-word.

   Multi-word conjunctions are labeled CONJP, with flat internal structure. Included in this list are "quasi-conjunctions", a somewhat open class of phrases that can function as conjunction or adverb, depending on context and annotator interpretation. See section 7.5.2 for more on the annotation of these phrases.

as well as
not to mention
rather than
instead of
if not (less frequently treated as a conjunction)
along with (less frequently treated as a conjunction)
etc.

```
(S (NP-SBJ That)
   (VP builds
       (NP (NP confidence)
           ,
           (NP self sufficiency)
           ,
           (CONJP not to mention)
           (NP critical regulatory net worth)))
   .)


(S (NP-SBJ She)
   (VP valued
       (NP wisdom
           (CONJP as well as)
           knowledge))
   .)
```

3. Discontinuous conjunctions.

   Only multi-word portions of discontinuous conjunctions are labeled CONJP. Single-word portions are left unlabeled.

not only ... but
not only ... but also
not only ... but instead
not only ... but rather
not alone ... but
not alone ... but also
not alone ... but instead
not alone ... but rather
not just ... but
not just ... but also
not just ... but instead
not ... but instead
not ... but
etc.

```
(S (NP-SBJ The proposal)
   (VP represents
       (NP (CONJP not alone)
           (NP his own district)
           but
           (NP (NP all the people)
               (PP of
                   (NP our country))))))
```

```
(S (NP-SBJ Her actions)
   (VP were
       (ADJP-PRD (CONJP not only)
                 (ADJP compassionate)
                 (CONJP but also)
                 (ADJP (ADJP inspiring)
                       and
                       (ADJP indicative
                             (PP of
                                 (SBAR-NOM what one can do
                                           with 50 cents)))))))
```

7.5.2 Quasi-coordinating conjunctions

The quasi-coordinating conjunctions are treated as ordinary coordinating conjunctions when they act as such. Otherwise, for example when fronted, they are bracketed as SBAR (e.g., *if not* and *though*) or PP (e.g., *than* and *as well as*).

if not
though
rather than
as well as
etc.

1. Fronted constituents. When a putative conjunction occurs as the first element of a fronted phrase, it is interpreted as the head of that phrase rather than as a conjunction, and the correct bracketing is ADVP or PP.

```
(S (PP Instead of
       (NP beans))
   ,
   (NP-SBJ I)
   (VP will
       (VP eat
           (NP pizza)))
   .)


(S (SBAR-ADV If not
             (FRAG-TMP next week))
   ,
   (NP-SBJ I)
   (VP suggest
       (SBAR 0
             (S (NP-SBJ you)
                (VP start
                    (NP-TMP (NP sometime)
                            (ADVP-TMP soon))))))
   .)
```

2. Adjuncts. When a putative conjunction occurs after the verb and introduces a sentential adjunct and where there is more than a typological difference between the putative conjuncts, the correct bracketing is ADVP or PP.

```
(S (NP-SBJ-2 I)
   (VP will
       (VP eat
           (NP take-out pizza)
           ,
           (PP instead of
               (S-NOM (NP-SBJ *-2)
                      (VP eating
                          (PP-LOC at
                                  (NP your house)))))))
   .)
```

3. Commas. The presence of a comma before expressions such as *rather than* and *instead of* sometimes suggests an ADVP or PP interpretation (depending on the judgment of the annotator). Thus, the following interpretations are equally likely:

```
(S (NP-SBJ I)
   (VP ate
       (NP pizza)
       ,
       (PP rather than
           (NP beans)))
   .)


(S (NP-SBJ I)
   (VP ate
       (NP (NP pizza)
           ,
           (CONJP rather than)
           (NP beans)))
   .)
```

Other conjunctions such as *as well as* retain the CONJP interpretation in the presence of a comma.

```
(S (NP-SBJ I)
   (VP ate
       (NP pizza
           ,
           (CONJP as well as)
           beans))
   .)
```

Note that some non-conjunction interpretations appear in the corpus due to a mid-corpus policy change. This section represents the most recent policy on labeling of competing conjunction/adverb interpretations and covers the majority of the annotation in the present corpus.

7.5.3 *times*

When it is used as a "quantifier", *times* is placed inside a QP (see the section on mathematical language in section 11 [Modification of NP]):

```
(NP (QP three times)
    the government's damages)
```

Otherwise, it retains the conjunction interpretation:

```
(S (NP-SBJ three times five)
   (VP equals
       (NP fifteen))
   .)
```

except in the case where it is preceded by the verb *multiply*, in which case the matter is unresolved. Fortunately, the Wall Street Journal corpus contains no such examples.

8 Shared Complements and Modifiers in Coordinate Structures

This section describes the general approach to the annotation of modifiers and complements that are shared by more than one head. Along the way, we will have to, of course, consider the bracketing of *unshared* modifiers; thus, much of the unshared-modifier policy is also described below. As the annotation of shared elements necessarily involves the annotation of coordinate structures, it may also be helpful to consult section 7 [Coordination].

Some other issues in coordination are also discussed here, including coordination of NP modifiers (see section 8.4) and certain complex NPs analyzed as having a shared head (see section 8.5).

8.1 Premodifiers

When it is not possible to tell from context whether or not a modifier is shared, the DEFAULT in all cases is to analyze the item as shared.

Shared bracketing in the case of NPs is flat.

In the case of shared premodifiers of VPs, the premodifier may be attached either immediately before the VP or immediately inside the VP. This variation occurs with both coordinated and non-coordinated VPs, and is regarded as semantically and syntactically insignificant.

```
(NP ripe apples and bananas)


(S (NP-SBJ Grover)
   (ADVP-MNR deliberately)
   (VP chewed and winked))


(S (NP-SBJ Grover)
   (VP (ADVP-MNR deliberately)
       chewed and winked))
```

Where it is clear that the modifier is not shared, we generally use more structure in order to indicate which head the modifier goes with:

```
(NP (NP ripe apples)
    and
    (NP cinnamon))


(NP fresh
    (NX (NX ripe apples)
        and
        (NX cinnamon)))
```

(See section 3 on page 140 and section 2 [Notation] for an explanation of the NX bracket label.)

```
(S (NP-SBJ Grover)
   (VP (VP (ADVP-MNR noisily)
           chewed)
       and
       (VP winked)))
```

In other respects, the annotation of premodifiers of verbs and that of nouns differ significantly, as laid out in the following sections.

8.1.1 Verbs

This section describes the annotation of verbs with shared adverbial premodifiers.

1. Labeling.

   All modifiers of VP are labeled, including one-word adverbs (this is in contrast to single-word modifiers of NP, which are left unlabeled).

   ```
   (S (NP-SBJ Curious George)
      (ADVP-MNR carefully)
      (VP cut and pasted
          (NP the pictures)))
   ```

   EXCEPTIONS to this labeling policy are *not* (whether sentential or constituent negation), single-word (including discontinuous) conjunctions such as *and, but, neither...nor* and floating quantifiers such as *all* and *both*. Some adverbs that behave somewhat like conjunctions are often also left unlabeled, such as *then, thus, only, so, also*.

   ```
   (S (NP-SBJ Curious George)
      (VP neither cut nor pasted
          (NP the pictures)))
   ```

2. Level of attachment.

   An adverbial that is outside the VP it modifies (either at coordination level or at S-level) is interpreted as shared. Again, the DEFAULT in unclear cases is to bracket the adverbial as shared.

   - Non-coordinated VPs.

     In the case of non-coordinated VPs, premodifiers are either left outside the VP, as in (a), or put inside it, as in (b). Although the variation is in general free, the adverbials inside the VP are considerably more likely to be -MNR or degree adverbs. Overall, the bracketing in (a) is more common.

     ```
     (a)  (S (NP-SBJ Curious George)
             (ADVP-MNR carefully)
             (VP cut
                 (NP the pictures)))

     (b)  (S (NP-SBJ Curious George)
             (VP (ADVP-MNR carefully)
                 cut
                 (NP the pictures)))
     ```

   - Coordinated VPs with shared modifiers.

     Shared premodifying adverbials follow the same tendencies for coordinated verbs as they do for single verbs, as described just above for "non-coordinated VPs".

```
(a)  (S (NP-SBJ Curious George)
        (ADVP-MNR carefully)
        (VP cut and pasted
            (NP the pictures)))


     (S (NP-SBJ Curious George)
        (ADVP-MNR carefully)
        (VP (VP cut
                (NP the pictures))
            and
            (VP pasted
                (NP them))))

(b)  (S (NP-SBJ Curious George)
        (VP (ADVP-MNR carefully)
            cut and pasted
            (NP the pictures)))


     (S (NP-SBJ Curious George)
        (VP (ADVP-MNR carefully)
            (VP cut
                (NP the pictures))
            and
            (VP pasted
                (NP them))))
```

- Note on unshared premodifiers in coordinate structures.
  In the case where a premodifier modifies only one of the conjuncts, the modifier is put inside the VP it modifies.

```
(S (NP Grover)
   (VP (VP (ADVP-MNR noisily)
           chewed)
       and
       (VP winked)))
```

There is, however, some variation in the bracketing of certain adverbs premodifying second VP conjuncts, particularly *then, thus, also, even, therefore, so,* which almost act like conjunctions. They are sometimes left at conjunction level, and they may be unlabeled or labeled ADVP.

```
(S (NP Grover)
   (VP (VP (ADVP-MNR noisily)
           chewed)
       and
       (ADVP then)
       (VP winked)))
```

8.1.2 Nouns

This section describes the annotation of pre-head modifiers of coordinated noun phrase heads, including adjectival and nominal modifiers, possessives, and determiners.

1. Nominal premodifiers.

   - Single head of NP.

     The interpretation of modifiers that are themselves nominal tends to be highly ambiguous and subject to individual interpretation. For example, in the noun phrase *the primary college entrance examination*, one person may have a clear intuition that the **college** is primary, while another may be sure that the **examination** is primary. Similarly, in *U.S. patent and copyright owners*, one person may think that the **owners** are U.S., while another may believe that the **patents and copyrights** are U.S.

     In order to avoid spending large amounts of time imposing arbitrary solutions to this problem, we try to avoid showing any structure for nominal modifiers:

     ```
     (NP the primary college entrance examination)

     (NP U.S. patent and copyright owners)

     (NP the loan and real estate reserves)
     ```

     In general, we avoid showing either the internal structure or the extent of modification of noun modifiers, regardless of the strength of the annotator's intuition in a particular example.

   - Multiple heads.

     This policy for NPs with single heads extends to those with multiple heads: if the only unshared modifiers are nominal, we annotate with flat structure.

     ```
     (NP the user and system identification or password)

     (NP Manhattan phone book and tour guide)

     (NP the new phone book and tour guide)

     (NP high-priced red Burgundies and Cabernets and Chardonnays)

     (NP tobacco consumption
         and
         lung-cancer mortality
         research)
     ```

     However, proper names are frequently annotated with internal structure (although officially they should be treated as other nominal modifiers).

     ```
     (NP (NP Arthur Dent)
         and
         (NP Ford Prefect))

     (NP (NP Mr. Kent)
         and
         (NP Ms. Lane))

     (NP (NP France)
         and
         (NP Hong Kong))
     ```

2. Other premodifiers.

   The interpretation of adjectives, possessives, and determiners tends to be more uniform, so more structure is shown for such modifiers when unshared.

   - Shared.

     When multiple heads shared the same modifiers, a flat structure is used.

```
(NP ripe apples and bananas)

(NP the apples and bananas)

(NP the seven ripe apples and bananas)

(NP your old apples and bananas)
```

However, the modifiers themselves get the same internal structure that they would get in a non-coordinated NP.

```
(NP the (ADJP very ripe) apples and bananas)

(NP (NP Sharon 's)
    apples and bananas)

(NP (NP (NP Sharon 's)
        and
        (NP Anthony 's))
    apples and bananas)


(NP seven
    (ADJP very old)
    apples and bananas)
```

- Unshared.

  When there are unshared modifiers, added structure (usually NP adjunction) shows which modifiers go with which head.

```
(NP (NP my dog)
    and
    (NP your cat))

(NP (NP apples)
    and
    (NP fresh basil))

(NP (NP (NP Bob 's)
        skirt)
    and
    (NP (NP Tracy 's)
        suit))
```

When there are unshared adjectives, determiners, or possessives, we frequently end up showing structure for nominal modifiers as well.

```
(NP (NP our science curriculum)
    and
    (NP our testing policies))


(NP (NP trade conflicts)
    and
    (NP sluggish exports))
```

```
(NP (NP accelerated unfair-trade investigations)
    and
    (NP stiff trade sanctions))

(NP-LGS (NP rising labor costs)
        and
        (NP the strong yen))
```

3. NX: combination of shared and unshared modifiers.

   So far we have only considered cases in which all the modifiers are shared or all the modifiers are unshared. When there is a mixture, we add a level of structure called NX.

   - **Definition of NX.**
     In the case where a noun is modified by an unshared modifier and also shares **non-nominal** premodifiers with another noun, the NX bracket label is used. That is, unshared items are lumped together with their respective noun heads and bracketed NX, with the shared modifier(s) outside NX (at NP level). The NX levels are then coordinated at the lowest level possible, as usual.
     NX brackets contain the head of the NP and its (unshared) modifiers in complicated NPs where both shared an unshared modifiers are involved. It does not correspond to any particular linguistic structure, although it occasionally resembles "N-bar". Rather, it exists only to show which modifiers go with which NP head, and is only used when the extent of modification would not otherwise be clear.

     The NAC label, described in section 11 [Modification of NP], plays a similar role of indicating modifiers that go together. However, NAC is only used for *pre-head modifiers*, while NX always contains the **head** (or heads) of the NP in which it is found.

   - Examples.

```
(NP some
    (ADJP very old)
    (NX (NX red apples)
        and
        (NX bananas)))

(NP the remarkable
    (NX (NX former New Democracy Party representative)
        and
        (NX (NX known political enemy)
            (PP of
                (NP Mr. Mitsotakis)))))

(NP (NP the farm stand 's)
    (NX (NX hard apples)
        and
        (NX mushy bananas)))
```

     Note that as before at NP level, nominal modifiers may end up with structure when one of the other conjuncts has an unshared modifier that is non-nominal.

```
(NP Manhattan
    (NX (NX phone book)
        and
```

```
        (NX exhausted guide)))

(NP the 187
    (NX (NX network affiliates)
        and
        (NX independent TV stations)))
```

Note that both conjuncts are labeled NX even in the case where only one of the conjuncts is multi-word:

```
(NP the
    (ADJP expensive and hard-to-find)
    (NX (NX ripe apples)
        and
        (NX cinnamon)))

(NP some
    (NX (NX red apples)
        and
        (NX bananas)))
```

- With nominal modifiers.

  When all of the unshared modifiers are nominal, NX should not be used. However, sometimes NX structure is used anyway, especially with proper nouns. But such structure is much less likely to happen when the coordination has to be at NX level than when it can be at NP level. (See page 138 on such structure at NP level.)

```
(NP the
    (NX (NX World Bank)
        and
        (NX International Monetary Fund)))
```

This may even happen when the **shared** modifiers are nominal, though that is even more unlikely.

```
(NP NY investors
    (NX (NX Douglas A. Kass)
        and
        (NX Anthony Pedore)))
```

4. (Pre)determiner vs. discontinuous conjunction.

   Note that some shared elements (e.g., *both, either*) have both (pre)determiner and conjunction uses. Roughly, the word in question receives the determiner analysis when it is referential and the conjunction analysis when it is part of a discontinuous conjunction.

   The policy with respect to the determiner/conjunction distinction assumed here follows the POS tagging policy given in the *POS guidelines* [Santorini 1990]. The distinction is made in the POS tagging but also has implications for the syntactic annotation.

   When single-word conjuncts are involved, the bracketing of conjunctions and shared determiners is identical (i.e., the annotation is flat). When it is not clear whether the determiner or conjunction analysis is correct, the DEFAULT is to annotate as conjunction.

   In the following examples, *both* and *either* are annotated as part of the discontinuous conjunctions *both...and* and *either...or*:

```
(NP both boys and girls)

(NP both digital and IBM systems)

(NP (UCP both
         Treebank
         and
         non-Treebank)
    policies)


(S (NP-SBJ Either
             (NP a boy)
             or
             (NP a girl))
   (VP could
       (VP sing)))


(S (NP-SBJ Either a boy or girl)
   (VP could
       (VP sing)))


(NP either sweet potato or mashed potato
    mix)
```

Compare with the following, where *both* and *either* are annotated as (pre)determiners:

```
(NP both
    (ADJP large, red and shiny)
    balls)


(NP both
    the
    (ADJP large, red and shiny)
    balls)


(S (NP-SBJ Either boy or girl)
   (VP could
       (VP sing)))


(NP either
    (NX (NX sweet potato
            (NX *RNR*-9))
        or
        (NX mashed potato
            (NX *RNR*-9))
        (NX-9 mix)))
```

8.2 Complements

8.2.1 Verbs

1. Overt complements.

   (a) Single-word VPs.

   The internal structure of coordinated single-word verbs is not shown and the shared object is attached under the VP.

```
(S (NP-SBJ John)
   (VP baked and frosted
       (NP the cake)))
```

   This is an example of what is meant by "coordinate low", where here coordination is at the level of V rather than VP. Note that the following structure is implicit in the annotation of single-word conjuncts and is retrievable from the POS tagging:

```
(S (NP-SBJ John)
   (VP (v (v baked)
          and
          (v frosted))
       (NP the cake)))
```

   (b) Multi-word VPs.

   "Multi-word VPs" may include negation, auxiliaries, particles, adverbs, other objects or adjuncts, etc. Shared complements in multi-word VPs are attached at the level of coordination and *RNR*-attached into each conjunct. See section 5 [Pseudo-Attach] for a description of *RNR*-attach.

```
(S (NP-SBJ John)
   (VP (VP baked
           (NP *RNR*-1))
       and
       (VP (ADVP-MNR carefully)
           frosted
           (NP *RNR*-1))
       (NP-1 the cake)))


(S (NP-SBJ John)
   (VP (VP likes
           (NP *RNR*-8))
       but
       (VP will
           not
           (VP buy
               (NP *RNR*-8)))
       (NP-8 the suit)))


(S (NP-SBJ Mary)
   (VP (VP handed
           (NP the suit)
           (PP-DTV *RNR*-6))
       and/but
```

```
        (VP mailed
            (NP the tie)
            (PP-DTV *RNR*-6))
        (PP-DTV-6 to
                  (NP John))))
```

The dictum **coordinate low** also requires that coordination be at the lowest possible VP level in the case where there are embedded VPs, as in the difference between *will have baked and may have frosted* and *will have baked and frosted*, where coordination is lower in the latter:

```
(S (NP-SBJ John)
   (VP (VP will
           (VP have
               (VP baked
                   (NP *RNR*-1)))
           and
           (VP may
               (VP have
                   (VP frosted
                       (NP *RNR*-1)))
               (NP-1 the cake)))))


(S (NP-SBJ John)
   (VP will
       (VP have
           (VP baked and frosted
               (NP the cake)))))


(S (NP-SBJ John)
   (VP will
       (VP (VP have
               (VP baked
                   (NP *RNR*-1)))
           and
           (VP have
               (VP frosted
                   (NP *RNR*-1)))
           (NP-1 the cake))))
```

If coordinated verbs share some complements (here, *spring goods*) but not others (here, *to Campeau stores*), the shared complement is *RNR*-attached into each conjunct and the stranded constituent is *ICH*-attached into the VP it is associated with.

```
(NP (NP the investment)
    (VP required
        (NP *)
        (S-PRP (NP-SBJ *)
               (VP to
                   (VP (VP make
                           (NP *RNR*-3))
                       and
                       (VP ship
                           (NP *RNR*-3)
```

```
                    (PP-DIR *ICH*-4))
                (NP-3 spring goods)
                (PP-DIR-4 to
                          (NP Campeau stores)))))))
```

Note: The term "multi-word VP" should not be confused with "multi-word verb", which refers to objects such as *spot check, pied pipe*, etc. The rare multi-word verbs in the corpus are bracketed flat, as follows with *test market*:

```
( (S (NP-SBJ Adolph Coors Co.)
     (VP said
         (SBAR 0
               (S (NP-SBJ its Coors Brewing Co. unit)
                  (VP will
                      (VP test market
                          (NP (NP a new line)
                              (PP of
                                  (NP bottled water)))
                          (PP-LOC in
                                  (NP the West))
                          (PP-TMP beginning
                                  (NP early next year)))))))
     .))
```

See section 26 [Orphans] for more on multi-word prepositions and multi-word adverbs.

(c) With auxiliaries.

Auxiliaries that share a verb are coordinated low (i.e., left flat), regardless of punctuation and possible intonation breaks that might otherwise suggest an *RNR* analysis.

```
(S (NP-SBJ I)
   (VP can and will
       (VP leave
           (PP-TMP by
                   (NP midnight)))))
```

A VP shared by a "multi-word auxiliary" (here, *soon will be*) is *RNR*-attached into each conjunct. (Note: Semi-auxiliaries such as *ought to* and *be able to* are bracketed as a VP — S series. See section 4 [Null Elements] for more on semi-auxiliaries.)

```
(NP (NP the number)
    (PP of
        (NP (NP country funds)
            (SBAR (WHNP-1 that)
                  (S (NP-SBJ-2 *T*-1)
                     (VP (VP are
                             (VP *RNR*-3))
                         or
                         (VP (ADVP-TMP soon)
                             will
                             (VP be
                                 (VP *RNR*-3)))
                         (VP-3 listed
                               (NP *-2)
```

```
                                        (PP-LOC in
                                              (NP (NP New York)
                                                  or
                                                  (NP London))))))))))
```

(d) Coordinated verbs that do not form a phrasal unit.

Some single word verbs may share an object without together forming a single VP. That is, in cases where an intonation break is required after the first verb, the structure will be annotated as coordinated VPs rather than coordinated Vs, with the shared argument *RNR*-attached to each single-word verb. Note that this *RNR*-attach is never available to single-word auxiliary verbs (see the preceding section "With auxiliaries").

```
(S (NP-SBJ the likely consequence)
   (VP would
       (VP be
           (S-PRD (NP-SBJ *)
                  (VP to
                      (VP (VP weaken
                              (NP *RNR*-1))
                          ,
                          (CONJP rather than)
                          (VP strengthen
                              (NP *RNR*-1))
                          (NP-1 (NP the control)
                                (SBAR that ...))))))))
```

It is recognized that the intonation break is a difficult and only marginally reliable test. So the DEFAULT in these cases is to coordinate low (i.e., leave as a single VP with coordinated Vs) unless the annotator feels quite sure that the *RNR* analysis is merited.

2. Null complements.

(a) Single-word VPs.

The null object in passive constructions (NP *) and the trace of wh-movement or topicalization (NP *T*) may also be shared. The trace is attached inside the flat VP. (See section 7 [Coordination] for the bracketing of coordinated single-word VPs.)

```
(NP (NP the athletes)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ they)
             (VP have
                 (VP wooed and won
                     (NP *T*-1))))))
```

In passive constructions, both the null, which is coindexed to the surface subject, and the *by*-phrase are arguments of VP attached inside VP.

```
(S (NP-SBJ-1 The pictures)
   (VP were
       (VP cut and pasted
           (NP *-1)
           (PP by
               (NP-LGS Curious George)))))
```

```
(S (NP-SBJ-1 The car)
   (VP was
       (VP washed and waxed
           (NP *-1))
       (PP by
           (NP-LGS Amy))))
```

(b) Multi-word VPs.

If one of the VPs consists of more than one word (*was washed and will be waxed*, *clipped and put on the refrigerator door*) or if the object must be attached at different levels (*was washed and sat on*, *see or look for*), then **multiple traces**, each with the same index, are used. That is, null elements are not *RNR*-attached and instead multiple instantiations of the null are shown.

- Passive examples.

```
(S (NP-SBJ-6 The car)
   (VP (VP was
           (VP washed
               (NP *-6)))
       and
       (VP will
           (VP be
               (VP waxed
                   (NP *-6))))))

(S (NP-SBJ-6 The car)
   (VP (VP was
           (VP washed
               (NP *-6)))
       and
       (VP polished
           (PRT up)
           (NP *-6))))

(S (NP-SBJ-6 The car)
   (VP (VP was
           (VP washed
               (NP *-6)))
       and
       (VP sat
           (PP-CLR on
                   (NP *-6)))))

(S (NP-SBJ-2 (NP (QP No fewer than 24)
                 country funds))
   (VP have
       (VP been
           (VP (VP launched
                   (NP *-2))
               or
               (VP registered
                   (NP *-2)
                   (PP-CLR with
                           (NP regulators)))
```

```
                    (NP-TMP this year)))))
```

- *Wh*-movement and topicalization examples.

```
(NP (NP one)
    (PP of
        (NP (NP those columns)
            (SBAR (WHNP-2 that)
                  (S (NP-SBJ you)
                     (VP (VP clipped
                             (NP *T*-2))
                         and
                         (VP put
                             (NP *T*-2)
                             (PP-PUT on
                                     (NP the refrigerator door)))))))))

((S (WHNP-1 Who)
    (SQ did
        (NP-SBJ you)
        (VP (VP see
                (NP *T*-1))
            or
            (VP look
                (PP-CLR for
                        (NP *T*-1)))))
    ?))

(S (NP-TPC-6 John)
   (NP-SBJ-7 I)
   (VP ca n't
       (VP stand
           (S (NP-SBJ *-7)
              (VP to
                  (VP (VP hear
                          (PP-CLR about
                                  (NP *T*-6)))
                      or
                      (VP see
                          (NP *T*-6))))))))
```

8.2.2 Nouns

(Note that in Treebank bracketing, only **clauses** (S or SBAR) are recognized as complements of NPs.)

1. Unshared complements.

   If the complement is unshared, it belongs inside only the relevant NP.

```
(NP (NP the basketball promise)
    and
    (NP his carefully thought-out decision
        (S to keep it)))
```

2. Shared complements.

   As with verbs, shared complements of nouns are placed at the level of coordination. (As usual, when it is not clear whether a complement is shared or not, the DEFAULT option is to bracket it as shared.)

```
(NP the belief and declaration
    (SBAR that the world is flat))

(NP group decisions and attempts
    (S to go to puppetry school))
```

   In cases where coordination does not result in flat structure (such as when there are unshared non-nominal premodifiers), the shared complement is placed at the level of coordination and *RNR*-attached into each conjunct.

```
(NP (NP his belief
        (SBAR *RNR*-5))
    and
    (NP your subsequent declaration
        (SBAR *RNR*-5))
    (SBAR-5 that
            (S the world is flat)))


(NP (NP the right
        (S *RNR*-1))
    , but not
    (NP the obligation
        (S *RNR*-1))
    ,
    (S-1 (NP-SBJ *)
         (VP to
             (VP sell
                 (NP a financial instrument)
                 (PP-CLR at
                         (NP a specified price))))))
```

   If modifiers force the use of NX, the complement clause should again be placed at the level of coordination and RNR-attached into each conjunct.

```
(NP the
    (NX (NX growing urge
            (S *RNR*-2))
        and
        (NX hard-to-resist temptation
            (S *RNR*-2))
        (S-2  (NP-SBJ *)
              (VP to
                  (VP sleep
                      (NP-TMP all morning))))))
```

8.2.3 Prepositions

1. Single-word PPs.

   Coordinated single-word PPs are annotated with flat structure just like other single-word conjuncts (e.g., *went in and out the door*).

```
(S (NP-SBJ The cat)
   (VP went
       (PP-DIR in and out
               (NP the door))))
```

2. Multi-word PPs.

   The term "multi-word PP" should not be confused with "multi-word preposition", which refers to sequences of prepositions that are annotated with flat structure (e.g., *instead of, because of*, etc.). (See section 26 [Orphans] for a list of multi-word prepositions).

   Multi-word PPs require an extra level of structure. Shared complements of multi-word PPs are attached at coordination level and *RNR*-attached into each conjunct.

```
(NP (NP A unit of data)
    (SBAR (WHNP-1 that)
          (S (NP-SBJ-2 *T*-1)
             (VP is
                 (VP moved
                     (NP *-2)
                     (PP-DIR (PP into
                                 (NP *RNR*-3))
                             or
                             (PP out
                                 (PP of
                                     (NP *RNR*-3)))
                             (NP-3 the computer)))))))
```

```
(S (NP-SBJ The average coupon)
   (VP is
       (PP-PRD (PP (NP-ADV (QP about 18) cents)
                   off
                   (NP *RNR*-3))
               ,
               or
               (PP (NP-ADV 15 percent)
                   off
                   (NP *RNR*-3))
               ,
               (NP-3 (NP the regular price)
                     (PP of
                         (NP the product))))))
```

   Such examples may be additionally annotated with PRN ("parenthetical") because of the commas, though the rest of the structure remains the same. See section 2 [Notation] for more on PRN.

```
(S (NP-SBJ The average coupon)
   (VP is
       (PP-PRD (PP (NP-ADV (QP about 18) cents)
                   off
                   (NP *RNR*-1))
               (PRN ,
                    or
                    (PP (NP-ADV 15 percent)
                        off
                        (NP *RNR*-1))
                    ,)
               (NP-1 (NP the regular price)
                     (PP of
                         (NP the product))))))
```

8.2.4 Adjectives

Adjectives are handled much like verbs. The shared constituent is attached at coordination level, and if the coordinated adjectives are multi-word, the shared item is *RNR*-attached.

```
(ADJP eager and ready
      (S (NP-SBJ *)
         (VP to
             (VP go))))


(ADJP (ADJP very eager
            (S *RNR*-5))
      but
      (ADJP (ADVP not quite)
            ready
            (S *RNR*-5))
      (S-5 (NP-SBJ *)
           (VP to
               (VP go))))
```

8.3 Adjuncts and postmodifiers

8.3.1 Verbs

1. Overt postmodifiers in the VP.

   - Flat VP → adjunct inside VP.

     Because all postverbal elements are attached inside the VP, postverbal shared modifiers of coordinated VPs are also put at coordination level.

```
(S (NP-SBJ The villain)
   (VP sang and danced
       (PP-LOC in
               (NP the park))))
```

- Multi-word VPs → attach at coordination level.

  If the coordination does not result in flat structure, the adjunct is placed at the lowest possible level of coordination, but not *RNR*-attached. Adjuncts attached at coordination level can be assumed to be interpreted at the same level in each conjunct, namely as an adjunct of each main verb.

```
(S (NP-SBJ-2 (NP (QP No fewer than 24)
                 country funds))
   (VP have
       (VP been
           (VP (VP launched
                   (NP *-2))
               or
               (VP registered
                   (NP *-2)
                   (PP-CLR with
                           (NP regulators)))
               (NP-TMP this year)))))


(S (NP-SBJ The villain)
   (VP (VP sang
           (ADVP-MNR brightly))
       and
       (VP danced
           (ADVP-MNR wildly))
       (PP-LOC in
               (NP the park))))
```

  The lowest possible level of coordination may be S-level:

```
(S (PP In
       (NP other words))
   ,
   (S (NP-SBJ economic growth)
      (VP would
          (VP be
              (ADJP-PRD lower))))
   and
   (S (NP-SBJ unemployment)
      (VP would
          (VP be
              (ADJP-PRD higher))))
   (PP-TMP for
           (NP a few years))
   .)
```

- Interpretation in different clauses → adjunct at coordination level, *RNR*-attach used.

  RNR-attach may be used for non-complements when they are shared across clauses. In the following example, the adjunct is extracted from a subordinate clause in the first conjunct but from a main clause in the second. *RNR*-attach is not normally used with adjuncts because adjuncts attached at coordination level can be assumed to be interpreted at the same level in each conjunct. In this case, however, simply attaching the locative adjunct at coordination level would give an incorrect interpretation (at the levels of *reported* and *injured*), and *RNR*-attach is necessary to achieve the correct interpretation (at the levels of *killed* and *injured*).

```
(S (S (NP-SBJ-3 (QP At least 270)
                people)
      (VP were
          (VP reported
              (S (NP-SBJ-1 *-3)
                 (VP-2 killed
                       (NP *-1)
                       (PP-LOC *RNR*-5))))))
   and
   (S (NP-SBJ=3-6 1,400)
      (VP=2 injured
            (NP *-6)
            (PP-LOC *RNR*-5)))
   (PP-LOC-5 in
             (NP (NP the rush-hour tremor)
                 (SBAR (WHNP-4 that)
                       (S (NP-SBJ *T*-4)
                          (VP caused
                              (NP (NP billions)
                                  (PP of
                                      (NP (NP dollars)
                                          (PP of
                                              (NP damage))))))))))))
```

2. Trace of wh-movement and topicalization.

   - Multi-word VPs → single adjunct trace at coordination level.

     The traces of shared adjuncts are put at the level of coordination, in the extraction site (i.e., just where an unmoved adjunct would be attached, see the preceding section, starting on page 151 on the attachment of shared adjuncts).

```
(S (NP-SBJ the bidding group)
   (VP has n't
       (VP had
           (NP (NP time)
               (SBAR (WHADVP-2 0)
                     (S (NP-SBJ *)
                        (VP (VP to
                                (VP develop
                                    (NP its latest idea)
                                    (ADVP-MNR fully)))
                            or
                            (VP to
                                (VP discuss
                                    (NP it)
                                    (PP-CLR with
                                            (NP banks))))
                            (ADVP-TMP *T*-2))))))))


(S (ADVP-TMP-TPC-1 Initially)
   ,
```

```
    (NP-SBJ the company)
    (VP said
        (SBAR 0
              (S (NP-SBJ-2 it)
                 (VP will
                     (VP (VP close
                             (NP its lending division))
                         , and
                         (VP stop
                             (S (NP-SBJ *-2)
                                (VP originating
                                    (NP new leases)
                                    (PP-LOC at
                                            (NP its lease subsidiary)))))
                         (ADVP-TMP *T*-1)))))))
```

Multiple traces are *not* used if they would all be attached at the same level (here, child of the first VP) in every VP conjunct. Rather, a single trace is attached at the lowest available coordination level.

Similarly, if a trace is shared by two Ss, it goes at coordination level.

```
(NP (NP a day)
    (SBAR (WHADVP-1 0)
          (S (S (NP-SBJ some United Airlines employees)
                (VP wanted
                    (S (NP-SBJ-2 Mr. Wolf)
                       (VP fired
                           (NP *-2)))))
             and
             (S (NP-SBJ takeover stock speculators)
                (VP wanted
                    (NP his scalp)))
             (ADVP-TMP *T*-1))))
```

- Interpretation in different clauses → multiple adjunct traces.

  Unlike with overt adjuncts in this structure (see page 152), *RNR*-attach is **never** used with traces.

  Multiple traces of adjunct *wh*-movement and topicalization are used only in the case that the adjunct is extracted from different clause levels (one matrix and one subordinate, for example), as this is the only structure where attaching the trace at coordination level will yield an incorrect interpretation. (See page 152.)

```
(NP (NP the town)
    (SBAR (WHADVP-4 where)
          (S (NP-SBJ the president)
             (VP (VP spoke
                     (ADVP-LOC *T*-4))
                 and
                 (VP declared
                     (SBAR that
                           (S (NP-SBJ-3 the building)
                              (VP should
                                  (VP be
```

```
                                                 (VP built
                                                     (NP *-3)
                                                     (ADVP-LOC *T*-4))))))))))
```

8.3.2 Nouns

See the Complements and Postmodifiers section of section 11 [Modification of NP] for more details on the bracketing of adjuncts and postmodifiers in NP.

Shared adjuncts are adjoined to the highest appropriate NP. *RNR*-attach (Right Node Raising) should not be used with adjuncts, although it is used with shared **complements** of nouns.

```
(NP (NP a book and poster)
    (PP about
        (NP toads)))


(NP (NP (NP a book)
        and
        (NP a poster))
    (PP about
        (NP reptiles)))


(NP (NP princes and dukes)
    (PP of
        (NP Luxemborg)))


(NP (NP (NP handsome princes)
        and
        (NP dignified dukes))
    (PP of
        (Luxemborg)))


(NP (NP the arrest and charging)
    (PP of
        (NP the two men)))


(NP-PRD (NP the
            (NX (NX crown prince)
                and
                (NX hereditary grand duke)))
        (PP of
            (NP Luxembourg)))
```

8.3.3 Comparative adjectives and adverbs

The *than/that/as*-clause in comparative structures is always adjoined to the comparative phrase. Thus, this type of postmodifier will be adjoined to comparative adjectives and adverbs. (See section 22 [Comparatives] for details on the bracketing of comparative structures.)

```
(ADJP (ADJP as long and complicated)
      (PP as
          (NP that paper)))

(ADJP (ADJP (ADJP as long)
            and
            (ADJP as complicated))
      (PP as
          (NP that paper)))

(ADVP (ADVP (ADVP as quickly)
            and
            (ADVP as efficiently))
      (PP as
          (ADJP possible)))

(ADJP (ADJP (ADJP less)
            (PP than
                (NP *RNR*-1)))
      or
      (ADJP equal
            (PP to
                (NP *RNR*-1)))
      (NP-1 the maximum link speed))
```

8.4 Coordination of adjectival and nominal NP modifiers

The structure of coordinated NP premodifiers is independent of the coordination of NP heads. Thus the rules below apply equally to NPs that have a single head or multiple, coordinated heads.

8.4.1 Adjectives

Conjoined single-word adjectives are labeled ADJP, with the internal structure left flat.

```
(NP (ADJP far-away and expensive)
    stores)

(NP (ADJP ripe and nutritious)
    apples and bananas)

(NP the
    (ADJP ripe and nutritious)
    apples and bananas)

(NP (NP two factors)
    ,
    (ADJP economic and political))
```

Conjoined multi-word adjectives are labeled ADJP, with the internal structure shown, even when just one of the conjuncts is multi-word.

```
(NP (ADJP (ADJP round)
          and
          (ADJP bright blue))
    balls)
```

```
(NP (ADJP (ADJP very large)
          and
          (ADJP extremely poisonous))
    apples)
```

```
(NP (NP a government)
    (ADJP (ADJP (ADJP less predictable)
                (PP than
                    (NP Mr. Gandhi 's)))
          , and
          (ADJP (ADVP possibly)
                more restrictive)))
```

When it is not clear whether the modifier of the adjective goes with just the first modifier or with both, it is assumed to be shared and the structure is left flat. This parallels the DEFAULT treatment of other shared elements when their scope is not clear from context:

```
(NP (ADJP very
          large and poisonous)
    apples)
```

Compare with the case where the scope of modification is shown:

```
(NP (ADJP (ADJP oddly smelling)
          and
          (ADJP poisonous))
    apples)
```

When a comma is used instead of a lexical conjunction, only multi-word adjectives are bracketed (unlike conjunction with *and*, *or*, etc., where both conjuncts are labeled).

```
(NP true , entertaining stories)
```

```
(NP smoother
    ,
    (ADJP less volatile)
    executions)
```

8.4.2 Nominal

No internal structure is shown for conjoined nominal premodifiers.

```
(NP installation and maintenance procedures)
```

```
(NP the installation and maintenance procedures)
```

```
(NP both installation and maintenance procedures)
```

```
(NP the human and animal health-products segment)
```

```
(NP the
    TV installation
    and
    antennae maintenance
    procedures)
```

Even in the case where a nominal premodifier is adjectivally modified, the entire structure is left flat.

```
(NP municipal bond
    and
    mutual fund
    orders)
```

However, nominal **post**modifiers get the same internal structure as other noun phrases.

```
(NP (NP Mickey Mouse)
    ,
    (NP editor and publisher))


(NP (NP Micky Mouse)
    ,
    (NP (NP treasurer)
        and
        (NP chief financial officer)))
```

8.4.3 Coordinated adjectival and nominal modifiers

When an adjective is coordinated with a nominal modifier, UCP ("unlike coordinated phrase") is used. (See section 7 [Coordination] for more information about UCP.)

No internal structure should be shown for a coordinated single-word noun and adjective premodifier:

```
(NP the
    (UCP federal and state)
    rulings and procedures)


(NP (NP sales)
    (PP-LOC at
            (NP (UCP franchisee
                     (CONJP as
                            well
                            as)
                     company-owned)
                stores)))
```

Multi-word conjuncts may show internal structure.

```
(NP (ADJP low-cost producing)
    (UCP (NP Pacific Rim)
         and
         (ADJP Latin American))
    countries)
```

8.5 Shared NP heads

NPs like *20 thin and 10 fat dogs*, in which unrelated modifiers are apparently conjoined, are analyzed as separate noun phrases sharing a common head. The common head is labeled NX, attached at conjunction level, and *RNR*-attached to each NP:

```
(NP (NP 20 thin
        (NX *RNR*-1))
    and
    (NP 10 fat
        (NX *RNR*-1))
    (NX-1 dogs))
```

Note that the shared NX is not limited to one word, and could in principle contain head coordination, although there are no actual examples of this in the WSJ corpus.

```
(NP (NP the Japanese
        (NX *RNR*-1))
    and
    (NP the U.S.
        (NX *RNR*-1))
    (NX-1 consumer markets))
```

This analysis extends in a natural (if somewhat ugly) way to unrelated modifiers that share premodifiers (such as a determiner).

```
(NP our
    (NX (NX 20 thin
            (NX *RNR*-1))
        and
        (NX 10 fat
            (NX *RNR*-1))
        (NX-1 dogs)))


(NP-SBJ (NP Gannett 's)
        (NX (NX 83 daily
                (NX *RNR*-1))
            and
            (NX 35 non-daily
                (NX *RNR*-1))
            (NX-1 newspapers)))


(NP-SBJ (NP Gannett 's)
        (NX (NX New York daily
                (NX *RNR*-1))
            and
            (NX Pennsylvania non-daily
                (NX *RNR*-1))
            (NX-1 newspapers)))
```

This *RNR*-attached NX analysis has also occasionally been used when a simpler UCP analysis would have sufficed:

```
(NP (NP municipal bond
        (NX *RNR*-2))
    ,
    (NP mutual fund
        (NX *RNR*-2))
    and
    (NP other
        (NX *RNR*-2))
    (NX-2 orders))
```

perhaps should be:

```
(NP (UCP municipal bond
         ,
         mutual fund
         and
         other)
    orders)
```

In fact, the NX analysis is not limited to simple conjunction of NPs. If the annotator has a strong intuition of Right Node Raising, an NX can be shared across a fairly complex structure:

```
(PP-LOC (PP in
            (NP normal
                (NX *RNR*-1)))
        (CONJP as well as)
        (PP in
            (NP cancerous
                (NX *RNR*-1)))
        (NX-1 cells))
```

Although the same construction is more likely to be bracketed more simply:

```
(PP-LOC (PP in
            (NP normal))
        (CONJP as well as)
        (PP in
            (NP cancerous cells)))
```

9 *WH*-phrases

This section is concerned with the bracketing of *wh*-phrases. For information about null elements associated with *wh*-phrases, see section 4 [Null Elements].

9.1 Bracketing *wh*-phrases in direct and indirect questions

9.1.1 Bracket labels

```
WHNP    WHADVP    WHADJP    WHPP
```

1. WHNP.

   (a) General.
   When *what, who*, and *which* stand alone, they are labeled WHNP.

   ```
   (SBARQ (WHNP-1 What/Who/Which)
          (SQ are
              (NP-SBJ you)
              (VP thinking
                  (PP-CLR about
                          (NP *T*-1)))))
          ?)
   ```

   When *how many* stands alone, it too is labeled WHNP. However, when it modifies a nominal head, it is bracketed WHADJP as noted in (c) below.

   ```
   (SBARQ (WHNP-9 How many)
          (S do
             (NP-SBJ you)
             (VP want
                 (NP *T*-9)))
          ?)
   ```

   (b) With modifiers.
   Single-word *wh*-premodifiers (e.g., *which* in *which dress*) are left unlabeled. The modifier and head are dominated by WHNP.

   ```
   (SBARQ (WHNP-1 What/Whose/Which dress)
          (SQ did
              (NP-SBJ you)
              (VP select
                  (NP *T*-1)))
          ?)
   ```

   Multi-word *wh*-premodifiers (e.g., *how many* in *how many secrets*) are bracketed according to the principle,

   Wh-ness percolates up.

   That is, the *wh*-word (here, *how*) and all higher-level nodes are bracketed with the appropriate *wh*-label. In most cases of premodification, the principle is moot; because we usually don't label the heads of NPs, there's no question as to whether they receive a *wh*-label.

```
(SBARQ (WHNP-8 (WHADJP how many)
               secrets)
       (SQ do
           (NP-SBJ you)
           (VP know
               (NP *T*-8))))
```

```
(SBARQ (WHNP-7 (WHADJP How hot)
               a room)
       (S can
          (NP-SBJ you)
          (VP tolerate
              (NP *T*-7)))
       ?)
```

The principle of upward percolation of *wh*-ness is more important to the bracketing of post-modified *wh*-phrases. The the *wh*-phrase *what, whose,* or *which* and all higher nodes are labeled WH[x], but nodes that do not dominate the *wh*-phrase are not labeled WH[x]. In the below examples, *in the closet* is labeled PP, not WHPP.

```
(S (NP-SBJ I)
   (VP do
       not
       (VP know
           (SBAR (WHNP-3 (WHNP what)
                         (PP-LOC in
                                 (NP the closet)))
                 (S (NP-SBJ I)
                    (VP am
                        (ADJP-PRD afraid
                                  (PP of
                                      (NP *T*-3)))))))))
   .)
```

```
(SBARQ (WHNP-1 (WHNP What/Whose/Which story)
               (PP about
                   (NP tribbles)))
       (SQ did
           (NP-SBJ you)
           (VP read
               (NP *T*-1)))
       ?)
```

2. WHADVP.

   (a) General.

   When *when, why, where,* and *how* stand alone they are simply labeled WHADVP. The *wh*-trace receives the appropriate function tag.

```
(S (SBAR-TMP (WHADVP-2 When)
             (S (NP-SBJ the clock)
                (VP strikes
                    (NP three)
                    (ADVP-TMP *T*-2))))
   ,
```

```
   (NP-SBJ the children)
   (VP leave)
   .)

(SBARQ (WHADVP-54 Why)
       (SQ did
           (NP-SBJ you)
           (VP jump
               (PP-DIR off
                       (NP the cliff))
               (ADVP-PRP *T*-54)))
       ?)

(SBARQ (WHADVP-1 Where)
       (SQ did
           (NP-SBJ you)
           (VP meet
               (NP them)
               (ADVP-LOC *T*-1)))
       ?)

(SBARQ (WHADVP-42 How)
       (SQ did
           (NP-SBJ you)
           (VP fix
               (NP the car)
               (ADVP-MNR *T*-42)))
       ?)
```

(b) With modifiers.

Single-word *wh*-premodifiers in WHADVPs (here, *how*) are left unlabeled, just as nominal single-word *wh*-premodifiers are.

```
(SBARQ (WHADVP-2 How fast)
       (SQ did
           (NP-SBJ the plants)
           (VP grow
               (ADVP *T*-2)))
       ?)
```

As with post-modified WHNPs, *wh*-ness percolates upward in postmodified WHADVPs. In this case, the prepositional phrase is labeled PP-LOC rather than WHPP-LOC.

```
(SBARQ (WHADVP-4 (WHADVP Where)
                 (PP-LOC in
                         (NP Minneapolis)))
       (SQ do
           (NP-SBJ you)
           (VP live
               (ADVP-LOC *T*-4)))
       ?)
```

(c) WHADJP

The WHADJP label is used to bracket phrases consisting of a *wh*-adverb modifier and an adjectival head.

```
(SBARQ (WHADJP-54 How cold)
       (SQ is
           (NP-SBJ it)
           (ADJP-PRD *T*-54)
           (ADVP-LOC outside))
       ?)
```

(d) WHPP

```
(SBARQ (WHPP-42 On
                (WHNP what))
       (SQ did
           (NP-SBJ you)
           (VP sit
               (PP-LOC-CLR *T*-42)))
       ?)
```

Note that premodified WHNP objects of WHPP are bracketed in accordance with the bracketing of all other premodified WHNPs.

```
(SBARQ (WHPP-6 In
               (WHNP which chair))
       (S did
          (NP-SBJ-5 you)
          (VP wish
              (S (NP-SBJ *-5)
                 (VP to
                     (VP sit
                         (PP-LOC *T*-6))))))
       ?)


(SBARQ (WHPP-2 In
               (WHNP (WHADJP how many)
                     chairs))
       (S did
          (NP-SBJ you)
          (VP sit
              (PP *T*-2)))
       ?)
```

9.1.2 Coordination of *wh*-phrases

1. Like *wh*-phrases are, as usual, coordinated as low as possible.

```
( (NP-SBJ I)
  (VP wonder
      (SBAR (WHNP-3 what or who)
            (S (NP-SBJ that)
               (VP is
                   (NP-PRD *T*-3)))))
  .)
```

There is no policy indicating how to bracket the trace of coordinated *wh*-phrases in cases where the conjuncts are associated with different function tags. Below, the issue is whether the (ADVP *T*) receives a -LOC, -TMP, or both. In some cases, the trace label contains no function tag, as in the first example. In others, the trace label contains all the appropriate function tags, as in the second example.

```
(S (NP-SBJ I)
   (VP forgot
       (SBAR (WHADVP-2 where and when)
             (S (NP-SBJ we)
                (VP will
                    (VP eat
                        (NP lunch)
                        (ADVP *T*-2))))))
   .)
```

```
(S (NP-SBJ I)
   (VP forgot
       (SBAR (WHADVP-2 where and when)
             (S (NP-SBJ we)
                (VP will
                    (VP eat
                        (NP lunch)
                        (ADVP-LOC-TMP *T*-2))))))
   .)
```

2. Unlike *wh*-phrases are dominated by UCP.

```
(S (NP-SBJ land developers)
   (VP tell
       (NP them)
       (SBAR (UCP-1 (WHADVP when)
                    ,
                    (WHADVP where)
                    ,
                    and
                    (WHPP in
                          (WHNP what manner)))
             (S (NP-SBJ the community)
                (VP shall
                    (VP grow
                        (UCP *T*-1)))))
       .))
```

The problem of how to bracket the *wh*-trace (noted in the preceding section on "Like *wh*-phrases") pertains here as well.

9.1.3 *ICH*-attaching to *wh*-phrases

In cases where a postposed constituent is interpreted as an adjunct to a *wh*-phrase, it is *ICH*-attached to the *wh*-phrase. See section 5 [Pseudo-Attach] for more information about *ICH*-attaching.

```
(SBAR-NOM (WHNP-2 (WHNP what)
                  (SBAR *ICH*-5))
          (S (NP-SBJ *T*-2)
             (VP is
                 (ADJP-LOC-PRD present)
                 (SBAR-5 (WHNP-1 that)
                         (S (NP-SBJ *T*-1)
                            (VP is
                                (VP creating
                                    (NP artificial volatility))))))))
```

9.1.4 Problematic cases

There are a number of tricky cases for which there is no policy. These occur infrequently in the present corpus. Possible bracketings are listed for each case.

1. *how come*; *how comes it*

   The question inversion appears to be tied up in the frozen form "how come"; hence the problem of SBAR vs SBARQ.

```
(SBAR[Q] how come
         (S[Q] (NP-SBJ you)
               (VP bushwhacked
                   (NP them rustlers))))
```

2. *what if*

```
(X (WHNP what)
   (SBAR if
         (S I told you pigs have wings)))
```

```
(SBAR (WHNP what)
      if
      (S...))
```

3. *why (not)*

```
(SBARQ (WHADVP why)
       (SQ (NP-SBJ *)
           not
           (VP grow
               (NP some)
               (PP (ADVP just)
                   for
                   (NP winter blooming)))))
```

```
(SBAR (WHADVP why)
      not
      (S ...))
```

```
(SBAR why not
      (S...))
```

4. *how('s) about*

```
(SBAR how about
      (S (NP-SBJ *)
         (VP watering
             (NP the plants)))
      ?)
```

5. *seeing (as how)*

```
(SBAR seeing as how
      (S (NP-SBJ the flowers)
         (VP need
             (NP attention))))
```

9.2 Bracketing *wh*-phrases in relative clauses

9.2.1 Bracket labels

1. WHNP

- Single-word WHNP.

```
(NP (NP answers)
    (SBAR (WHNP-6 that/which)
          (S (NP-SBJ-3 we)
             (VP 'd
                 (VP like
                     (S (NP-SBJ *-3)
                        (VP to
                            (VP have
                                (NP *T*-6)))))))))
```

- Premodified WHNP.

```
(NP (NP the teacher)
    (SBAR (WHNP-1 whose scarf)
          (S (NP-SBJ I)
             (VP admired
                 (NP *T*-1)))))
```

- Postmodified WHNP.

  In most cases, these cases are bracketed according to the principle "*wh*-ness percolates upward." However, since this principle was applied somewhat inconsistently, deviations as shown in the second example are likely.

```
(NP (NP three blind mice)
    (SBAR (WHNP-4 (NP none)
                  (WHPP of
                        (WHNP which)))
          (S (NP-SBJ-1 I)
             (VP was
```

```
                    (VP (ADVP very)
                        charmed
                        (NP *-1)
                        (PP by
                            (NP-LGS *T*-4)))))))
```

```
(SBAR (WHNP-107 (WHNP some)
                (WHPP of
                      (WHNP whom)))
      (S (NP-SBJ *T*-107)
         (VP do n't
             (VP have
                 (NP adequate staffs)))))
```

2. WHADVP

```
(NP (NP the place)
    (SBAR (WHADVP-2 that/where)
          (S (NP-SBJ I)
             (VP put
                 (NP the book)
                 (ADVP-PUT *T*-2)))))
```

```
(NP (NP the time)
    (SBAR (WHADVP-1 when)
          (S (NP-SBJ I)
             (VP met
                 (NP you)
                 (ADVP-TMP *T*-1)))))
```

Note that the SBARs do not have -LOC or -TMP labels. Only the adverbial trace receives a function tag.

3. WHPP

```
(NP (NP the place)
    (SBAR (WHPP-1 in
                  (WHNP which))
          (S (NP-SBJ I)
             (VP matured
                 (PP-LOC *T*-1)))))
```

```
(NP (NP the woman)
    (SBAR (WHPP-1 of
                  (WHNP which))
          (S (NP-SBJ I)
             (VP am
                 (ADJP-PRD fond
                           (PP *T*-1))))))
```

9.2.2 Null *wh*-elements/zero relatives

For information about bracketing null *wh*-elements, see section 4 [Null Elements].

9.2.3 Free ("headless") relatives

1. General. A free or headless relative is defined as any relative clause that lacks a head. Free ("headless") relatives are labeled SBAR-NOM.

```
(PP instead of
    (S-NOM (NP-SBJ *)
           (VP listening
               (PP-CLR to
                       (SBAR-NOM (WHNP-155 what)
                                 (S (NP-SBJ *T*-155)
                                    (VP is
                                        (PP-LOC-PRD in
                                                    (NP his soul))))))))


( (S (SBAR-NOM-SBJ (WHNP-1 What)
                   (S (NP-SBJ *T*-1)
                      (VP is
                          (PP-PRD of
                                  (NP (NP (ADJP much more)
                                          importance)
                                      (PP to
                                          (NP the Colombian economy))
                                      (PP than
                                          (NP (NP the supposed
                                                  benefits)
                                              (PP of
                                                  (NP laundered drug
                                                      money)))))))))
     (VP is
         (NP-PRD (NP higher prices)
                 (PP for
                     (NP (NP Colombia 's)
                         legitimate products))))
     .))
```

2. Distinguishing between free relatives and indirect questions.

   When the SBAR is a complement of PP or in subject position, it is clearly a free relative, and thus labeled SBAR-NOM.

   When the SBAR is a complement of a VP, it may be interpreted as either a free relative or a clausal complement.

   In the first example (below), the SBAR is bracketed as a clausal complement (in this case, an indirect question) and does not receive the -NOM tag. The sentence can be paraphrased as , "I asked, what did he ask?"

   In the second example, the clause is bracketed as a free relative and does receive the -NOM tag. The sentence can be paraphrased as "I asked that which he asked" or "I asked the same question that he asked."

- indirect question interpretation

```
(S (NP-SBJ I)
   (VP asked
       (SBAR (WHNP-1 what)
             (S (NP-SBJ he)
                (VP asked
                    (NP *T*-1)))))
   .)
```

- free-relative interpretation

```
(S (NP-SBJ I)
   (VP asked
       (SBAR-NOM (WHNP-1 what)
                 (S (NP-SBJ he)
                    (VP asked
                        (NP *T*-1)))))
   .)
```

SBAR complements of verbs such as *ask, tell,* and *know* that can take clausal complements are usually analyzed as clausal complements. Free-relative interpretations of clausal complements of VP happen infrequently, and are either due to a bona-fide free-relative interpretation, or to error.

```
(NP (NP the first)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *T*-1)
             (VP to
                 (VP tell
                     (NP him)
                     (SBAR (WHNP-154 what)
                           (S (NP-SBJ *T*-154)
                              (VP is
                                  (PP-LOC-PRD in
                                              (NP our minds))))))))))
```

9.3 Long-distance movement

Instances of "long-distance movement" (i.e., the *wh*-phrase is interpreted in a clause that is more deeply embedded than the free relative clause itself) are bracketed as follows.

- free relative

```
( (S The company is upset over
     (NP (NP the use)
         (PP of
             (SBAR-NOM what
                       (S (NP-SBJ it)
                          (VP says
                              (SBAR (WHNP-1 0)
                                    (S (NP-SBJ *T*-1)
                                       (VP are
                                           (NP-PRD its exclusive
```

```
                                                          trademarks)))))))))
          .))
```

- indirect question

```
( (S (NP-SBJ I)
     (VP forgot
         (SBAR (WHNP who)
               (S (NP-SBJ they)
                  (VP said
                      (SBAR (WHNP-2 0)
                            (S (NP-SBJ-1 they)
                               (VP wanted
                                   (S (NP-SBJ *-1)
                                      (VP to
                                          (VP hire
                                              (NP *T*-2))))))))))))
     .))
```

10 Subordinate Clauses

10.1 Scope of this chapter

Of the various kinds of subordinate clauses, this section is concerned only with those that are introduced by subordinating conjunctions. See section 1 [Overview of Basic Clause Structure] for information about sentential adjunct clauses, section 15 [Small Clauses] for small clauses, and section 9 [*WH*-Phrases] for subordinate clauses introduced by *wh*-phrases.

10.2 Definition of subordinating conjunction

The Treebank brackets as subordinating conjunctions those constructs which introduce finite clauses, past participle clauses, and sentence fragments. Most subordinate present participle clauses are classified as nominal gerunds and introduced by prepositions (with a few exceptions, such as clauses introduced by *while*). See section 13 [Gerunds and Participles] for more information about the bracketing of present participles.

Note that many words can function both as subordinating conjunctions and prepositions. A word is considered a subordinating conjunction (heading an SBAR) when it introduces a sentence, and a preposition (heading a PP) when it introduces nominals and other complements.

10.3 Distribution of subordinating conjunctions

10.3.1 Sentential/verbal adjunct

```
(S (SBAR-TMP After
             (S (NP-SBJ she)
                (VP finished)))
   ,
   (NP-SBJ she)
   (VP took
       (NP a shower)))


(S (NP-SBJ-1 She)
   (VP read
       (NP the paper)
       (SBAR-TMP while
                 (S (NP-SBJ *-1)
                    (VP eating
                        (NP breakfast))))))
```

10.3.2 Adjunct or complement of noun

```
(S (NP-SBJ (NP Her muscles)
           (SBAR-TMP before
                     (S (NP-SBJ-1 she)
                        (VP started
                            (S (NP-SBJ *-1)
                               (VP lifting
                                   (NP weights)))))))
   (VP were
```

```
        (ADJP-PRD smaller)))

  (NP the belief
      (SBAR that
            (S (NP-SBJ the world)
               (VP is
                   (ADJP-PRD round)))))

  (NP (NP such a pretty butterfly)
      (SBAR that
            (S (NP-SBJ I)
               (VP smiled
                   (PP-CLR with (NP delight))))))
```

When a postmodifying clause is introduced by *as*, there is usually no *wh*-element.

```
  (NP (NP the scenario)
      (SBAR as
            (S (NP-SBJ-4 *)
               (VP depicted
                   (NP *-4)
                   (PP by
                       (NP-LGS the middle-of-the-road group))))))
```

10.3.3 Predicate

```
  (S (NP-SBJ The perfect time)
     (VP is
         (SBAR-PRD after
                   (S (NP-SBJ she)
                      (VP finishes)))))
```

10.3.4 Complement of VP

```
  (S (NP-SBJ Willie)
     (VP knew
         (SBAR (WHNP-1 who)
               (S (NP-SBJ *T*-1)
                  (VP threw
                      (NP the ball))))))
```

10.3.5 Object of PP

assuming (that)
excepted (that)
excepting (that)
given (that)
granted (that)
granting (that)
provided (that)

providing (that)
save (that)
seeing (that)
supposing (that)
etc.

```
(S (PP Given
       (SBAR that
             (S (NP-SBJ raindrops)
                (VP are
                    (ADJP-PRD blue)))))
   ,
   (NP-SBJ (NP it)
           (S *EXP*-2))
   (VP does n't
       (VP make
           (NP sense)
           (S-2 to get a blue umbrella)
           (S-PRP because they will just blend in))))
```

10.3.6 *so ... that*

```
(S (NP-SBJ These beliefs)
   (VP (ADVP (ADVP so)
             (SBAR *ICH*-5))
       dominate
       (NP (NP our educational establishment)
           ,
           (NP our media)
           ,
           (NP our politicians)
           ,
           and
           (NP even
               our parents))
       (SBAR-5 that
               (S (NP-SBJ (NP it)
                          (S *EXP*-1))
                  (VP seems
                      (ADJP almost
                            blasphemous))
                  (S-1 (NP-SBJ *)
                       (VP to
                           (VP challenge
                               (NP them))))))))
```

10.4 SBARs in comparative constructions

See section 22 [Comparatives] for information about this subject.

10.5 "Absolute *with*" constructions

In line with the general policy stated above, present participle clauses are bracketed S-NOM and dominated by PP and past participle clauses are bracketed S and dominated by SBAR. See section 13 [Gerunds and Participles] for more information.

```
(SBAR with
      (S (NP-SBJ-3 the flowers)
         (VP arranged
             (NP *-3)
             (PP by
                 (NP-LGS Penelope)))))


(PP with
    (S-NOM (NP-SBJ her eyes)
           (VP rolling)))
```

When *with* is followed by coordinated NPs/S-NOMs and past participles, policy is undetermined. The following disparate bracketings are possible.

```
(S (NP-SBJ the end result)
   (VP would
       (VP be
           (NP-PRD (NP a leaner , meaner corporate America)
                   ,
                   (PP with
                       (UCP (NP soaring productivity and profits)
                            and
                            (S (NP-SBJ the weaker)
                               (VP gone
                                   (PP-DIR to
                                           (NP the wall)))))))))


(S (SBAR-ADV With
             (S (S (NP-SBJ demand)
                   (VP growing))
                and
                (S (NP-SBJ workers)
                   (PP-PRD in
                           (NP short supply)))))
   ,
   (NP-SBJ many Japanese manufacturers)
   (VP are
       (VP spending
           (ADVP-MNR heavily)
           (PP-CLR on
                   (NP automation)))))
```

10.6 Bracketing of subordinating conjunctions

10.6.1 Single-word

The following words are single-word subordinating conjunctions and are bracketed SBAR.

after, although, as
because, before
for
if
like
once
since, so
than, though
unless, until
whether, while

```
(SBAR though
     (S the world is no longer flat))

(SBAR though
     (FRAG (ADJP flat)))
```

10.6.2 Multi-word

The following words are bracketed SBAR. No internal structure is shown.

as if
as though
in case
in order to/that/for
in that
inasmuch as
insofar as
so as
so that
such that
whether or not

```
(SBAR as if
     (S he had walked for miles))
```

Note that *such that* is not always a subordinating conjunction:

- Subordinating conjunction:

```
(S (NP-SBJ X)
   (VP divides
       (PP-CLR into Y)
       ,
       (SBAR-ADV such that
                 (S (NP-SBJ the result)
```

```
                    (VP is
                        (NP Z))))))
```

- Not a subordinating conjunction:

```
(S (NP-SBJ the weather)
   (VP is
       (ADJP-PRD such
                 (SBAR that
                       (S you ought to bring your umbrella)))))
```

Note that *or not* in *whether or not* is attached unlabeled in SBAR whether the S precedes it or follows it.

```
(S (NP-SBJ I)
   (VP don't
       (VP know
           (SBAR whether or not
                 (S I should buy this)))))

(S (NP-SBJ I)
   (VP don't
       (VP know
           (SBAR whether
                 (S I should buy this)
                 or not))))
```

The collocation *now that* is most often bracketed:

```
(ADVP-TMP now
          (SBAR that
                (S ...)))
```

10.6.3 Modified subordinating conjunctions

1. Degree/extent modifiers. These are, for the most part, left unlabeled. However, in some cases (most often, with *especially*), they have been labelled ADVP. Examples of both bracketings are given below.

   only
   even
   just
   especially

```
(SBAR only
      because
      (S you agree to pay that 500 dollars))


(SBAR (ADVP only)
      because
      (S you agree to pay that 500 dollars))
```

2. Quantitative modifiers.

   The following are equally likely:

```
(SBAR-TMP (NP-ADV two weeks)
          before
          (S (NP-SBJ they)
             (VP departed)))


(SBAR-TMP (NP two weeks)
          before
          (S (NP-SBJ they)
             (VP departed)))
```

10.7 Correlative *the*-clauses (*the...the...* constructions)

There is no definitive policy for handling these cases. Most analyses involve the use of SBAR. See section 25 [Correlative *the*-Clauses] for more on the bracketing of correlative *the*-clauses (*the...the* constructions).

```
(S (SBAR-ADV (X the sooner)
             (S our vans hit the road each morning))
   ,
   (X the easier)
   it is for us to fulfill that obligation)
```

11 Modification of NP

The policies described in this section apply to cases in which the modifier is not shared by coordinated heads. See section 8 [Shared Complements and Modifiers] for information on the annotation of both coordinated heads and coordinated modifiers.

11.1 Premodifiers

11.1.1 Adjectives

1. Single-word ADJPs are not labeled.

```
(NP poisonous apple)
```

2. Multi-word ADJPs are labeled ADJP.

```
(NP (ADJP very poisonous) apple)

(NP (ADJP nearly invalid) license)

(NP (ADJP Holliston , Mass.-based) company)

(NP (ADJP well to do) people)
```

3. Hyphenated adjectives are considered single-word (since they form single word-tokens in the POS tagging) and are therefore not labeled.

```
(NP pre-historic apple)

(NP a Massachusetts-based company)
```

11.1.2 Nominal modifiers

Since it is often impossible to determine the scope of nominal modifiers, they are not labeled.

```
(NP fake sales license)

(NP white-water rafting license)

(NP week-end sales license)

(NP furniture sales license)

(NP fake fur sale)

(NP State Secretary inauguration)

(NP New York public officials)
```

Likewise, titles that precede proper names are not labeled:

```
(NP club president Helen Parker)

(NP State Secretary James Baker)
```

Note, however, that nominal modifiers containing PPs are not left flat. Instead, the PP is fully annotated and the nominal modifier is labeled NAC ("Not A Constituent"). The NAC structure here does not indicate complementation, despite appearances to the contrary, and is to be considered on a par with other cases of post-nominal adjunction. (See section 11.2.1.)

```
(NP (NAC sale
         (PP of
             (NP firecrackers)))
    law)


(NP (NAC Secretary
         (PP of
             (NP State)))
    James Baker)
```

NP premodifiers of words such as *ago* and *before* are labeled NP. See note at the end of section 5 on page 190.

The coordination of nominal modifiers is not annotated:

```
(NP installation
    and
    (NAC maintenance
         (PP of
             (NP software)))
    procedures)


(NP (NAC installation
         (PP of
             (NP hardware)))
    and
    (NAC maintenance
         (PP of
             (NP software)))
    procedures)


(NP (NAC installation & maintenance
         (PP of
             (NP software)))
    procedures)
```

11.1.3 Possessives

The possessive marker (usually *'s*, but sometimes just an apostrophe) is treated as an individual token – it is separated from the previous word and part-of-speech tagged POS. (However, possessive pronouns (including *its*) are treated as a single token, part-of-speech tagged PP$.)

We indicate what is doing the possessing by annotating the possessor as a noun phrase, attaching the possessive marker as the last child of the noun phrase.

```
(NP (NP my best friend 's)
    boyfriend)
```

A possessive *'s* phrase is **always** labeled NP, even if the possessor is single-word (because the possessive marker is a separate token).

```
(NP (NP Sharon 's)
    bananas)
```

Complicated possessive phrases are handled by using the usual rules for noun phrases, and then attaching the possessive marker as the last child of the possessing noun phrase. Hence, the possessing NP can be analyzed by removing the possessive marker and analyzing the remainder in the same way as for ordinary NPs. (The NAC label, which would be used in the below case if the possessive marker was not present, should **not** be used for possessives.)

```
(NP (NP (NP First)
        (PP of
            (NP America))
        's)
    operating results)
```

Possessives can be nested:

```
(NP (NP (NP Reader 's)
        Digest Association 's)
    new Magazine Publishing Group)
```

or serial:

```
(NP (NP China 's)
    (NP People 's)
    Daily)
```

Possessive phrases can also sometimes function as nouns:

```
(S (NP-SBJ It)
   (VP was
       (NP-PRD (NP my decision)
               , not
               (NP anyone else 's))))
```

```
(S (NP-SBJ Kim 's)
   (VP has
       (NP yummy food)))
```

11.1.4 Dates, places, expressions of amount

Nominal modifiers that are expressions of measure or amount, dates, or places are treated as adjectives in the simple cases:

```
(NP a five-dollar book)
```

```
(NP a 379-245 vote)
```

```
(NP June 30, 1989)
```

```
(NP the Jan. 12 meeting)

(NP the New York meeting)
```

More complex cases may involve the labels QP, NAC, etc. See section 11.3 for more details on the annotation of complex measure and amount phrases, dates, and places in NP.

```
(NP a
    (ADJP $ 5-a-share *U*)
    increase)


(NP a
    (ADJP (QP 10 to 15) lb.)
    monkey)


(NP a
    (ADJP (QP $ 200 million) *U*)
    contract)


(NP the
    (NAC-TMP Jan. 12, 1984)
    meeting)


(NP my
    (NAC-LOC New York
             ,
             NY)
    birthplace)
```

11.1.5 Substantive adjectives

Substantive adjectives are labeled NP. An elided head noun is not represented overtly, but is nonetheless recoverable from the annotation: if the last child of a base NP is an ADJP or JJx, then that adjective is either the head of the NP or modifying a null head (depending on one's theory of substantives).

```
(NP the rich)

(NP the best)

(NP (NP the best)
    (PP of
        (NP friends)))
```

Coordinated substantive adjectives or substantive adjectives that are modified (i.e., multi-word ADJPs) are labeled ADJP within a headless NP. Again, an elided head noun is not represented in the annotation.

```
(NP the (ADJP best and brightest))

(NP the
    (ADJP (ADJP very best)
          and
          (ADJP most talented)))
```

The internal structure of substantive ADJPs follows the same rules as for other ADJPs. See section 8 [Shared Complements and Modifiers] for more on coordination and the annotation of shared modifiers.

11.1.6 Participial and gerund modifiers

When it is not clear whether a modifier is an adjective/participle or a noun/gerund, annotators refer to the POS tag or the tests listed in the *POS guidelines* [Santorini 1990]. (See section 13 [Gerunds and Participles] for more details on the annotation of participles and gerunds acting as heads.)

1. Participles.

   Participial modifiers are bracketed like adjectival modifiers. Therefore, the head of an ADJP may be POS-tagged VBN or VBG.

```
(NP a
    flying
    plane)

(NP a
    (ADJP flying and competing)
    plane)

(NP A
    (ADJP Swiftly Tilting)
    Planet)

(NP a
    (ADJP professionally flying & competing)
    plane)

(NP (ADJP publicly traded)
    portfolios)
```

2. Gerunds.

   Gerund modifiers are bracketed like nominal modifiers.

```
(NP a
    baking
    guidebook)

(NP a
    baking and frosting
    guidebook)

(NP a
    vegetarian cooking
    guidebook)
```

11.2 Complements and Postmodifiers

11.2.1 Adjuncts

1. General

   (a) All postmodifiers are Chomsky-adjoined to the phrase they modify, with the exception of clausal complements of certain nouns (e.g, deverbal nouns). See section 11.2.2 for information on distinguishing complements from relative clauses.

```
(NP (NP the books)
    (PP-LOC on
            (NP the shelf)))

(NP (NP books)
    (PP of
        (NP prayer)))

(NP (NP writers)
    (ADJP full
          (PP of
              (NP promise))))

(NP (NP the earthquake)
    (NP-TMP yesterday))

(NP (NP women)
    (SBAR (WHNP-1 who)
          (S (NP-SBJ *T*-1)
             (VP love
                 (NP potato chips)))))
```

   Exception: Only **clausal complements** of NP are placed inside NP.

```
(NP the belief
    (SBAR that the world is round))
```

   (b) One could argue that the structures in the previous section do not reflect the true structure of noun phrases. For example, *the lamp which is near the window* can't be understood really as

```
(NP (NP the lamp)
    (SBAR which is near the window))
```

   because *the lamp* can't refer successfully; the *the* means that the rest of the phrase must pick out a unique lamp, which it doesn't. So the "right" structure must be something like:

```
(NP the
    (??? lamp
         (SBAR which is near the window)))
```

   However, making this kind of distinction on a regular basis would make our NP structure too complicated to be annotated at reasonable speed, so we settle for the simplified structures shown in the previous section on page 184.

This policy also applies to complex cases in which it is more obvious that our structure doesn't represent the "truth". For example, (ii) may better reflect the annotator's understanding of the text (as well as the correct underlying structure), in that it is *the government's damages* which are undetermined, rather than *three times the government's damages* as implied by (i). Nonetheless, (i) shows the correct annotation.

```
(i)  (NP (NP (QP three times)
             (NP the government 's)
             damages)
         ,
         (SBAR (WHNP-1 which)
               (S (NP *T*-1)
                  (VP are
                      (ADVP-TMP presently)
                      (ADJP-PRD undetermined)))))


(ii)  (NP (QP three times)
          (??? (NP the government 's)
               damages
               ,
               (SBAR (WHNP-2 which)
                     (S (NP-SBJ *T*-2)
                        (VP are
                            (ADVP-TMP presently)
                            (ADJP-PRD undetermined))))))
```

(c) -TMP and -LOC are the only tags used for NP postmodifiers. For instance, the NP in a *by*-phrase should not receive a -LGS (logical subject) tag.

```
(NP (NP the destruction)
    (PP of
        (NP Rome))
    (PP by
        (NP me)))
```

Contrast with the case where the *by*-phrase is a postmodifier of *destroy*, a verb:

```
(NP (NP the city)
    (VP destroyed
        (NP *)
        (PP by
            (NP-LGS me))))
```

See section 1 [Overview of Basic Clause Structure] and section 2 [Notation] for more details on the use of the -LGS tag.

(d) *ICH*-attach is never used to show word order of adjuncts within NPs. (i) shows the correct treatment of displaced adjuncts within NP; (ii) shows an incorrect use of *ICH*-attachment with adjuncts.

```
(i)  (NP-SBJ (NP A report)
             (NP-TMP late yesterday)
             (SBAR (WHNP-2 that)
                   (S (NP-SBJ I)
```

```
                    (VP liked
                        (NP *T*-2)))))

(ii)  (NP-SBJ (NP A report)
              (SBAR *ICH*-3)
              (NP-TMP late yesterday)
              (SBAR-3 (WHNP-2 that)
                      (S (NP-SBJ I)
                         (VP liked
                             (NP *T*-2)))))
```

Note that *ICH*-attach is used within nouns only when other postmodifiers intervene between the head noun and its **complement**:

```
(NP (NP the decision
        (S *ICH*-1))
    (PP-LOC in
            (NP upper management))
    (S-1 (NP-SBJ *)
         (VP to
             (VP hire
                 (NP new people)))))
```

2. Reduced relative clauses

   - VP-based.

     When the postmodifier contains an overt VP, but no subject and no subordinating conjunction, all items are attached inside the VP and the VP is adjoined to the NP. See section 13 [Gerunds and Participles] for more details on the annotation of reduced relative clauses.

```
(NP (NP women)
    (VP loving
        (NP women)))

(NP (NP women)
    (VP loving
        (NP women)
        (PP-TMP in
                (NP this century))))
```

Note that passive traces are not coindexed when they occur in reduced relative clauses. This reflects an understanding of the relationship between the NP and reduced relative as post-modification rather than predication. See section 4 [Null Elements] for more information about the annotation of reduced relative clauses.

```
(NP (NP meals)
    (VP enjoyed
        (NP *)
        (PP by
            (NP-LGS everybody))))
```

- Non-VP-based.

  Annotation of reduced relative clauses that do not contain VPs varies as follows. When there are several postmodifying elements, possibly all part of the same relative clause, the preferred approach is to analyze only one adjunct as part of the relative clause, leaving the rest to be attached at S or VP level (as appropriate) if possible, as in (a).

```
(a)  (S (NP-SBJ I)
        (VP read
            (NP (NP the books)
                (PP-LOC on the shelf))
            (NP-TMP yesterday)))
```

  In the case where such an interpretation is impossible, as in *I read the books on the shelf yesterday quickly and the books on the shelf today slowly*, where *yesterday* must go inside the NP and cannot be attached at a higher level, official policy is to use the RRC label, as in (b), though more likely annotations are as listed in (c), where the postnominal elements are adjoined non-recursively, and (d), where the postnominal elements are placed within a single constituent which is then adjoined to the NP. See section 13 [Gerunds and Participles] for more information about non-VP-based reduced relatives.

```
(b)  (VP read
         (NP (NP the books)
             (RRC (PP on (NP the shelf))
                  (NP-TMP yesterday))))


(c)  (NP (NP the books)
         (PP-LOC on
                 (NP the shelf))
         (NP-TMP yesterday))


(d)  (NP (NP 110 titles)
         (PP-LOC not
                 (ADVP-TMP presently)
                 in
                 (NP the collection)))
```

3. Non-recursivity of adjuncts.

   (a) Consecutive unrelated adjuncts are non-recursively attached to the NP they modify. Relative clauses are also non-recursively attached to the NP containing the head noun.

```
(NP (NP the woman)
    (PP-LOC in the store)
    (SBAR who sold me the book))


(NP (NP State University)
    (PP of NY)
    (PP-LOC at Albany))


(NP (NP the book)
    (SBAR I read yesterday)
    ,
    (SBAR which is called Bread and Jam...))
```

(b) Exceptions to the the non-recursivity policy in the previous section on page 187.

- Apposition. Appositives are recursively adjoined to the NP they modify.
  That is, an NP dominates everything before the comma, an NP dominates everything after the comma, and the two are adjoined under a higher NP, leaving intact the structure of other postmodifiers of the NP or appositive.

```
(NP (NP (NP the book)
        (SBAR I read yesterday))
    ,
    (NP a historical novel))
```

  Note that nonrestrictive relative clauses are not considered appositives and are therefore not recursively adjoined:

```
(NP (NP the book)
    (SBAR I read yesterday)
    ,
    (SBAR which is called Bread and Jam...))
```

- Postmodifying phrases following the expressions *a share, per share* are recursively adjoined. (Note that this is not the case with similar expressions such as *a day, per person*, etc, after which subsequent postmodifying phrases are nonrecursively adjoined.)
  The expressions *a share* and *per share* may in some cases be treated differently: The phrase *a share* is always adjoined to the head noun when the NP containing the head noun immediately precedes it, as in (a).

```
(a)  (NP (NP (NP 5 dollars)
             (NP-ADV a share))
         (PP in interest))
```

  However, there is some variation in the annotation of postmodifying *per share* when it immediately follows the NP containing the head noun, where the NP may contain recursive adjunction (b) or not (c).

```
(b)  (NP (NP (NP 5 dollars)
             (PP per share))
         (PP in interest))


(c)  (NP (NP 5 dollars)
         (PP per share)
         (PP in interest))
```

  When there is intervening material, *a share/per share* is not recursively adjoined.

```
(NP (NP 5 dollars)
    (PP in interest)
    (NP-ADV a share))
```

(c) -TTL.
Constituents tagged -TTL (title) are bracketed recursively, regardless of whether it is the head NP or the appositive that bears the -TTL tag.

```
(NP (NP (NP the book)
        (SBAR I read yesterday))
    ,
    ``
```

```
        (NP-TTL (NP Bread and Jam)
                (PP for
                    (NP Frances)))
        '')

    (NP ``
        (NP-TTL (NP Bread and Jam)
                (PP for
                    (NP Frances)))
        ''
        ,
        (NP (NP the book)
            (SBAR I read yesterday)))
```

4. Reflexive pronouns.

   Reflexives are adjoined to the NP they follow.

```
(NP (NP the boys and girls)
    (NP themselves))


(NP (NP he)
    (NP himself))
```

   Note that if the reflexive falls elsewhere in the sentence, it is not pseudo-attached to the noun and is instead labeled NP-ADV.

```
((S (NP-SBJ He)
    (VP did
        (NP it)
        (NP-ADV himself))
    .))
```

5. *Alone, else, much, all.*

   Postmodifiers such as *alone, else,* and *much* are for the most part adjoined to the NP they follow; occasionally no internal structure is shown. There is also some variation as to whether the word is labeled ADVP or ADJP. The hoped-for, most common bracketings are shown below.

```
(NP (NP its real estate)
    (ADVP alone))

(NP (NP they)
    (ADVP alone))

(NP (NP anything)
    (ADJP much))
```

   When a quantifier immediately follows an NP, as in *they all*, it is usually bracketed as follows:

```
(NP (NP they)
    (NP all))
```

The bracketing of *else* has the most variation; all three possibilities are likely.

```
(NP (NP someone)
    (ADJP else))

(NP (NP someone)
    (ADVP else))

(NP someone else)
```

Note: Words such as *ago* and *before*, which are easily mistaken for postmodifiers, are instead annotated as the head of a phase that takes an NP premodifier. See section 11.3 for more on the annotation of measure and amount phrases.

```
(ADVP-TMP (NP weeks)
          ago)

(ADVP-TMP (NP two weeks)
          before)
```

To distinguish a premodifying NP from the NP complement of the PP, the premodifier is usually tagged -ADV. However, due to annotator variation, about a third of such NPs lack the -ADV. It role is recoverable from the structure, however.

```
(PP-TMP (NP-ADV two weeks)
        before
        (NP their departure))

(PP-TMP (NP two weeks)
        before
        (NP their departure))
```

Similar variation exists for NP premodifiers of subordinate clauses:

```
(SBAR-TMP (NP-ADV two weeks)
          before
          (S they departed))

(SBAR-TMP (NP two weeks)
          before
          (S they departed))
```

11.2.2 Clausal complements

1. Bracketing.

   (a) Clausal complements of nouns are placed inside NP as follows.

   ```
   (NP the desire
       (S to dance wildly on the roof))


   (NP the fact
       (SBAR that she wants that particular book))
   ```

   (b) The sentential complement of a noun is *ICH*-attached to the head NP if other postmodifiers intervene.

   ```
   (NP (NP the decision
           (S *ICH*-1))
       (PP-LOC in
               (NP upper management))
       (S-1 (NP-SBJ *)
            (VP to
                (VP hire
                    (NP new people)))))
   ```

2. Distinguishing clausal complements from relative clauses.

   An S or SBAR is bracketed as a complement when it follows certain nouns (e.g., deverbal nouns) and/or when it and the associated noun can be paraphrased as a subject-predicate pair.

   (a) Following is a partial list of words which take complements.

   S : *desire permit proposal option temptation authority contract negotiations attempt chance decision power right ability*

   SBAR : *fact idea proposal claim*

   (b) Examples of noun/complement → subject/predicate paraphrases:

   S : *the desire to dance wildly on the roof* →
   The desire is *to dance wildly on the roof*

   SBAR : *the fact that the young girl was courageous* →
   The fact is *that the young girl was courageous.*

   (c) An S is analyzed as a relative clause when

      i. the associated noun is not on the above wordlist, and the S can't be paraphrased as part of subject/predicate pair, but **can** be paraphrased with a *wh*-phrase.
      time to go →
      time **at which** *to go*

      ii. it has a "gap", i.e. a place an NP or *wh*-phrase can be interpreted.
      time at which to go **then**

   (d) Note that some of the nouns on the above complement-taking list may be followed by a clause which is paraphrasable with a *wh*-phrase. This enables annotators to bracket them as taking a relative clause when context suggests doing so.

   For example, *authority* usually takes a complement. But in the NP below, the SBAR is shared by *authority* and *funds*, which does not take a clausal complement. There is, therefore, a *wh*-trace in the SBAR. Since this trace would be inappropriate in a complement clause, *authority* is instead analyzed as taking a relative clause (paraphrasable as *authority* **under which** *to build fallout shelters...*).

```
(NP (NP authority
        and
        funds)
    (SBAR (WHADVP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP build
                     (NP (NP fallout shelters)
                         (VP costing
                             (NP (QP about 200 million)
                                 dollars)))
                     (ADVP-CLR *T*-1))))))
```

11.2.3 Reduced relative vs. floating participle

See section 13 [Gerunds and Participles] for a lengthy discussion of this distinction.

1. Reduced relatives.

   Reduced relatives are identified as having the following properties: they are (i) strongly associated with a noun, (ii) may not be paraphrased with *while* or *being*, and (iii) may be paraphrased with *which is* or *who is*.

   ```
   (S (NP-SBJ (NP The progress)
              (VP reported
                  (NP *)
                  (PP by
                      (NP-LGS the advisory committee))))
      (VP is
          (ADJP-PRD real)))
   ```

2. Floating participles.

   Floating participles are identified as having the following properties: they (i) may be moved around the sentence without fundamentally changing the relationship of the participle to the sentence, (ii) may be paraphrased with *while* and *being*, and (iii) may not be paraphrased with *which is* or *who is*. See section 13 [Gerunds and Participles] for more on the annotation of participles.

   ```
   (S And
      (ADVP-TMP now)
      ,
      (NP-SBJ-1 the woman)
      ,
      (S-ADV (NP-SBJ *-1)
             (UCP (ADJP-PRD tired)
                  and
                  (VP trembling)))
      ,
      (VP came
          (ADVP-DIR here)
          (PP-DIR to
                  (NP the DeKalb County cannery))))
   ```

3. Variation.

   The application of the above tests varies from annotator to annotator and sometimes different tests will result in different bracketing. For example, the following are both likely interpretations of the *replacing...* phrase.

   - floating participle

```
(S (NP-SBJ-1 (NP The Rusk belief)
             (PP in
                 (NP balanced defense)))
   ,
   (S-ADV (NP-SBJ *-1)
          (VP replacing
              (NP (NP the Dulles theory)
                  (PP of
                      (NP massive retaliation)))))
   ,
   (VP removes
       (NP a grave danger)))
```

   - reduced relative

```
(S (NP-SBJ-1 (NP (NP The Rusk belief)
                 (PP in
                     (NP balanced defense)))
             ,
             (VP replacing
                 (NP (NP the Dulles theory)
                     (PP of
                         (NP massive retaliation))))
             ,)
   (VP removes
       (NP a grave danger)))
```

11.3 Measure/Amount Phrases

11.3.1 QP (quantifier phrase)

This label is **not** used for NPs with quantificational determiners such as *every, some, almost all,* etc. Instead, it is used for multiword **numerical** expressions that occur within NP (and sometimes ADJP), where the QP corresponds frequently to some kind of complex determiner phrase.

The determiners *a* and *an* are included in the QP in cases where the appropriate interpretation is *one*:

```
(NP (QP under an) hour)
```

```
(NP (QP less than a) year)
```

When expressions such as *from...to...* form a complex determiner, they are labeled QP.

```
(NP (QP from 10 to 15)
    monkeys)
```

```
(NP (QP 10 to 15)
    monkeys)
```

```
(NP (QP 85 to 90 million)
    lbs.)
```

```
(NP (QP 10 or 11 million)
    lbs.)
```

```
(NP (QP between 12 and 13)
    percent)
```

```
(NP (QP between
        four and five
        hundred thousand)
    humans)
```

Note that there are instances (as described in the next section 11.3.2) where *from...to...* is treated as a PP rather than QP, corresponding roughly to cases where the measure phrase is a post-nominal modifier. Post-nominal modifiers are distinguished from **postposed** modifiers such as the following, in which the *from...to...* phrase is labeled QP:

```
(NP (NP gorillas)
    (NP (QP from 800 to 1000)
        lbs.))
```

```
(NP (NP a number)
    (PP between
        (NP 0 and 6)))
```

```
(PP between
    (NP (NP 5 dollars)
        and
        (NP 10 dollars)))
```

Discontinuous QPs are annotated as follows, where *or more* is considered to be part of the QP *two or three*:

```
(NP (QP two or three)
    inches
    (QP or more))
```

In the following examples, *an* and *800* are the first part of a discontinuous QP, but being single words are not labeled as such, in accordance with general policy.

```
(PP-TMP in
        (NP an hour
            (QP or so)))
```

```
(NP (NP gorillas)
    (NP 800 lbs.
        (QP and up)))
```

In cases where the head noun is missing (understood from context), we place the labeled QP inside an otherwise empty NP bracket. This applies only to multi-token numbers and their modifiers or to multi-word non-numeric "quantifiers." Examples are common in stock quotes, where "points" is often implied but rarely written out (see IBM example below).

```
(NP (QP as many as 10))
```

- single token numbers with no numeric modifiers (unlabeled)

  average circulation of 4,393,237

```
(NP (NP average circulation)
    (PP of
        (NP 4,393,237)))
```

  an additional 243,677 of the Class C warrants

```
(NP (NP an additional 243,677)
    (PP of
        (NP the Class C warrants)))
```

- multi-token numbers (and their modifiers):

  IBM rose 3 5/8 [points]

```
(NP-EXT (QP 3 5/8))
```

  8 million [cars] broke down on the freeway last year

```
(NP (QP 8 million))
```

  This car seats more than 5 [people]

```
(NP (QP more than 5))
```

- multi-word non-numeric "quantifiers":

  I paid more than double [the original price]

```
(NP (QP more than double))
```

  Hey, you ate more than half [the pie]

```
(NP (QP more than half))
```

- When a measure such as "pound" is morphologically singular, it is labeled ADJP rather than NP. This type of ADJP may also contain a QP expression.

```
(NP a
    (ADJP (QP 10 to 15)
          lb.)
    monkey)
```

```
(NP a
    (ADJP $ 5-a-share *U*)
    increase)
```

11.3.2 Ranges and endpoints: *from...to...*

Where a range is indicated, *from...to...* is annotated as a complex (conjoined) PP; where two end points are indicated, *from* and *to* are annotated as separate (nonconjoined) PPs. The distinction is made using the following test: if the order of the PPs in question can be reversed, then they constitute endpoints, and if not, they constitute a range. Note that *from...to...* ranges in determiner position are called QP, as in the example *from 10 to 15 monkeys* above, on page 193. The following examples contain nouns modified by ranges/endpoints.

- range:

```
(NP (NP a number)
    (PP (PP from
            (NP 2))
        (PP to
            (NP 32))))
```

```
(NP (NP excursions)
    (PP (PP from
            (NP studio))
        (PP to
            (NP studio))))
```

- endpoints:

```
(NP (NP the transition)
    (PP from
        (NP vinyl records))
    (PP to
        (NP compact discs)))
```

- range or endpoints, depending on context:

```
(PP from
    (NP June 15))
```

```
(PP to
    (NP June 30 , 1989))
```

When ranges or endpoints modify a verb, the picture is much the same as with nouns:

- range:

```
(VP varied
    (PP (PP from
            (NP 30))
        (PP to
            (NP 53 mg.))))
```

- endpoints:

```
(VP went
    (PP-DIR from
            (NP Paris))
    (PP-DIR to
            (NP Dakar)))


(VP went
    (PP-DIR from
            (NP general))
    (PP-DIR to
            (NP specific terms)))
```

11.3.3 Symbols in the text

1. Units of measure: *U*

   In cases where the head noun of a measure phrase appears as a symbol (such as $ or %) whose position precludes its being bracketed as the head noun, *U* is inserted as a place-holder meaning "[unit]" or "[units]".

   ```
   (NP (QP $ 200 million) *U*)

   (NP a
       (ADJP (QP $ 200 million) *U*)
       contract)
   ```

   When the QP is a single word, it is not labeled. Symbols such as $ in the example below are not counted as "words" in making the single- or multiple-word distinction:

   ```
   (NP $ 5 *U*)
   ```

   In cases where a symbol **can** be bracketed as the head noun, *U* is unnecessary:

   ```
   (NP (QP between 12 to 13) %)
   ```

This contrasts with cases where the symbol in question appears with both numbers, where *U* is required:

```
(NP (QP between 12 % to 13 %) *U*)

(NP (NP (QP between $ 150 and $ 200) *U*)
    (NP-ADV a week))
```

The above policy works much better with "dollars" and $ than with "cents", which appears as a word rather than a symbol throughout the corpus. As a result, some variation exists in the bracketing of examples with ranges of *cents*.

For example, *eight cents to 10 cents* may be bracketed in the following ways:

```
(NP (NP eight cents)
    to
    (NP 10 cents))


(NP (QP eight cents to 10) cents)

(NP (QP eight cents to 10 cents) *U*)
```

2. Mathematical language

   plus and *times* are bracketed as conjunctions:

```
(NP (NP (ADJP old and new)
        photos)
    plus
    (NP a written statement))


(S (NP-SBJ three
           times
           four)
   (VP is
       (NP-PRD twelve)))
```

11.3.4 Measure phrases in other syntactic environments

1. Without *of*

   Measure phrases without *of* are bracketed as adjunction structures:

```
(NP (NP one tablespoon)
    (NP quick-cooking tapioca))
```

2. With prepositions and adverbs

   Measure phrases that modify prepositions or adverbs are placed inside the phrase and are labeled NP (and not QP).

```
(ADVP-TMP (NP two weeks)
          ago)

(ADVP-TMP (NP two weeks)
          before)

(PP-TMP (NP two weeks)
        before
        (NP their departure))


(ADVP-TMP (NP an hour
              (QP or so))
          later)


(VP bury
    (NP him)
    (PP-LOC (NP six feet)
            under
            (NP the ground)))
```

Compare with the case where the measure phrase does not modify the preposition, but rather the PP modifies the NP.

```
(PP-TMP during
        (NP (NP the two weeks)
            (PP-TMP before
                    (NP their departure))))
```

3. Height and width

```
(ADJP (NP 3 ft.)
      long)

(NP (NP 3 ft.)
    (PP in
        (NP length)))


(NP (NP (QP from 20 to 60)
        feet)
    (PP in
        (NP height)))


(PP at
    (ADJP (ADVP just under)
          (ADJP (NP four feet)
                tall)
          and
          (ADJP (NP two feet)
                wide)))
```

Note that *just under* here is annotated as an adverbial modifier of the ADJP, much like *approximately*, etc.

4. Scores

   Scores are left with flat structure. They are labeled ADVP when they occur with verbs.

```
(NP a 379-245 vote)

(S (NP-SBJ they)
   (VP won
       (ADVP 97-94)))


(S (NP-SBJ they)
   (VP won
       (ADVP 97 to 94)))

(S (NP-SBJ The Knicks)
   (VP destroyed
       (NP Orlando)
       (ADVP 137-82)))
```

11.3.5 Multipliers: *times, half as much,* etc.

Expressions of amount such as *times, half as much,* etc. are labeled QP when they can be analyzed as complex determiners.

```
(NP (NP (QP three times)
        (NP the government 's)
        damages)
    ,
    (SBAR (WHNP-1 which)
          (S (NP-SBJ *T*-1)
             (VP are
                 (ADVP-TMP presently)
                 (ADJP-PRD undetermined)))))


(S (NP-SBJ BLAH)
   (VP (VP occupies
           (NP (NP (QP half as much)
                   floor space)
               (PP as
                   (NP older systems))))
       but
       (VP can
           (VP store
               (NP (QP five times as much)
                   data)))))


(VP bought
    (NP (QP more than double)
        that amount))
```

```
(NP (NP (QP nine times as many)
        frogs)
    (PP as
        (NP toads)))
```

When similar expressions modify a verb, they are labeled ADVP rather than QP.

```
(VP (ADVP more than)
    double
    (NP its purchases))
```

11.3.6 An alphabetized Bestiary of treatments of measure and quantifier phrases

(American Heritage Dictionary (1991): "A medieval collection of allegorical fables about the habits and traits of animals, each fable followed by an interpretation of its moral significance")

- *about*

```
(NP (QP about 1000)
    people)
```

- *all but*

```
(NP (QP all but 4)
    states)
```

  (When *all but* modifies a verb, in the sense of "did everything except", it is annotated as a flat ADVP.)

- *and up*

```
(NP (NP gorillas)
    (NP 800 lbs.
        (QP and up)))
```

- *around*

```
(NP (QP around 1000)
    people)
```

- *as many as*

```
(NP (QP as many as 15)
    names)
```

```
(NP (NP (QP as many as five million))
    (PP of
        (NP its common shares)))
```

- *as much as*

```
(NP (NP (QP as much as 15) %)
    (PP of
        (NP Jaguar shares)))
```

There may be occasional irregularities in the treatment of *as much as*, where it appears with the bracketing shown below, which is consistent with the usual structure for comparatives but inconsistent with just about everything else:

```
(NP (NP as much)
    (PP as
        (NP 15 %)))
    (PP of
        (NP Jaguar shares)))
```

- *at least*

```
(NP (QP at least 5)
    people)
```

- *from X on...*

```
(PP (PP from
        (NP there))
    (ADVP on))
```

```
(VP begins
    (PP (PP from
            (NP number 100))
        (ADVP on)))
```

- *just under*

Note that *just under* here is annotated as an adverbial modifier of the ADJP, much like *approximately*, etc.

```
(PP at
    (ADJP (ADVP just under)
          (ADJP (NP four feet)
                tall)
          and
          (ADJP (NP two feet)
                wide)))
```

- *more than*

```
(NP (QP more than one)
    person)
```

```
(NP (QP more than one)
    chimpanzee)

(NP (QP more than three in five))

(NP (QP 1, 2, 3, 4, and more than 4)
    orangutans)

(NP (NP a value)
    (NP (QP no more than 8)
        characters))


(NP (NP (ADJP (NP (QP 15 % to 30 %) *U*)
              more)
        output)
    (PP than
        (NP the current crop)))


(S (NP-SBJ I)
   (VP want
       (NP (NP more)
           (PP than
               (NP money)))))
```

- *nearly*

```
(NP (QP nearly 1000)
    people)
```

- *only*

```
(NP (QP only 1000)
    people)
```

- *out of*

```
(NP (NP (QP three out of five)
        skilled workers)
    and
    (NP (QP one out of five)
        technicians))
```

- *over*

```
(NP (QP over 1000)
    people)
```

- *through*

```
(NP (NP a number)
    (PP (PP from
            (NP 0))
        (PP through
            (NP 6))))
```

As a general rule, when *through* is not in construction with another preposition and occurs between two like categories, such as NPs, it is annotated as a conjunction:

```
(NP numbers
    4 through 9)

(NP (NP the file mode number)
    (PRN *LRB*
         (NP 0 through 6)
         *RRB*))
```

- *under*

```
(NP (QP under 1000)
    people)
```

- *up to*

```
(NP (QP up to 1024)
    bites)

(NP (NP (QP up to $ 15,000) *U*)
    (NP-ADV a month))
```

- *upwards of*

```
(NP (QP upwards of 1000)
    people)
```

11.4 Dates and places

The annotations of dates and places are parallel in many respects.

11.4.1 Dates

Dates are labeled NP when they are not adjectival modifiers of some other NP. They may or may not receive the adverbial -TMP tag depending on their function in the sentence. The internal structure of the date NP is left flat.

```
(NP June 30, 1989)
```

```
(NP (NP the meeting)
    (PP-TMP on
            (NP March 8, 1924)))


(NP (NP (NP Bill 's)
        birthday)
    ,
    (NP-TMP April 12, 2001))
```

Dates that are adjectival modifiers inside an NP are bracketed in one of two ways. If they contain a comma or *the*, they are labeled NAC ("Not A Constituent") and given the -TMP tag. Internal structure of NAC is not shown:

```
(NP the
    (NAC-TMP Jan. 12, 1984)
    meeting)


(NP the
    (NAC-TMP Friday the 13th)
    stock plunge)
```

Otherwise, they are left flat:

```
(NP the Jan. 12 meeting)
```

11.4.2 Places

Place-phrases are labeled NP when they are not adjectivally modifying some other NP. They may or may not receive the adverbial -LOC tag depending on their function in the sentence. The internal structure of the place NP is annotated with adjoined structure.

```
(NP (NP IBM)
    ,
    (NP-LOC (NP Akron)
            ,
            (NP OH)))


(NP (NP IBM)
    ,
    (VP based
        (NP *)
        (PP-LOC-CLR in
                    (NP (NP Akron)
                        ,
                        (NP OH))))
    ,)


(NP (NP the meeting)
    (PP-LOC in
            (NP (NP New York)
                ,
```

```
                (NP NY))))
```

Places-phrases that modify NPs are bracketed in one of two ways. If they contain a comma or are otherwise "complex", they are labeled NAC ("Not A Constituent") and given the -LOC tag. The internal structure of NAC is not shown.

```
(NP the
    (NAC-LOC New York
             ,
             NY
             ,)
    meeting)


(NP (NP Bill 's)
    (NAC-LOC Newark
             ,
             NJ)
    birthplace)


(NP my
    (NAC-LOC New York
             ,
             NY)
    birthplace)
```

Otherwise, they are left flat:

```
(NP the New York meeting)
```

11.5 Proper nouns

Proper nouns are bracketed in the same way as common nouns. There are no special rules concerning them.

```
(NP Free Press financial statements)

(NP the Free Press)

(NP Knight-Ridder officials)

(NP Xerox marketing strategies)

(NP the
    (NX (NX Free Press)
        and
        (NX New York Public Library)))
```

Since nominal modifiers of nouns are usually left flat, proper noun modifiers should also be left flat.

```
(NP the Free Press and New York Public Library scandal)
```

12 Titles

12.1 Whole constituent as title

Only titles of books, movies, songs, and names of other created works are labeled -TTL. (Note that -TTL implies -NOM, so no constituent need be tagged both -TTL and -NOM.)

```
(S ‘‘
   (NP-TTL-SBJ (NP Bedtime)
               (PP for
                   (NP Frances)))
   ’’
   (VP is
       (NP-PRD my favorite book)))


(S (NP-SBJ I)
   (VP like
       (S-TTL (NP-SBJ *)
              (VP Driving
                  (NP Miss Daisy)))))


(S (SBAR-TTL-SBJ (WHADVP-2 When)
                 (S (NP-SBJ Harry)
                    (VP Met
                        (NP Sally)
                        (ADVP-TMP *T*-2))))
   (VP was
       (NP-PRD a good movie)))


(S (NP-SBJ I)
   (VP am
       (PP-PRD in
               (NP (NP awe)
                   (PP of
                       (S-TTL (NP-SBJ The Empire)
                              (VP Strikes
                                  (ADVP-CLR Back))))))))


(NP ‘‘
    (PP-TTL In
            (NP (NP the Heat)
                (PP of
                    (NP the Night))))
    ,
    ’’
    (NP the NBC series...))
```

Names of institutions, games, and companies should not be labeled -TTL.

```
(PP according
    (PP to
        (NP (NP the Center)
```

```
            (PP for
                (NP (NP Continuing Study)
                    (PP of
                        (NP the California Economy)))))))
```

The label at the level of coordination involving items with -TTL is NP, without an additional -TTL.

```
(NP (NP-TTL Crime and Punishment)
    ,
    (S-TTL (NP-SBJ One)
           (VP Flew
               (PP-DIR Over
                       (NP (NP the Cuckoo 's)
                           Nest))))
    and
    (NP-TTL (NP The House)
            (PP-LOC at
                    (NP Pooh Corner))))
```

12.2 Premodified titles

When a title that would otherwise constitute a full NP is preceded by a modifier or determiner, it is labeled NX to create a placeholder for the -TTL tag.

```
(NP the uptempo
    ``
    (NX-TTL Sky)
    '')


(NP (NP Saint-Saens 's)
    ``
    (NX-TTL The Swan)
    '')


(NP (NP Steinbeck 's)
    ``
    (NX-TTL (NX The Grapes)
            (PP of
                (NP Wrath))))
```

However, some variation exists with postmodified NXs, especially with NX-TTL. Some such examples in the corpus are bracketed as follows:

```
(NP (NP Steinbeck 's)
    ``
    (NX-TTL (NP The Grapes)
            (PP of
                (NP Wrath))))
```

12.3 Titles as premodifiers

A complex constituent acting as a nominal modifier may be annotated with full structure and tagged -TTL.

```
(NP (NP the late Martin Luther King 's)
    famous ``
    (S-TTL (NP-SBJ I)
           (VP Have
               (NP a Dream)))
    '' speech)


(NP the premiere
    ``
    (PP-TTL In
            (NP the Dumpster))
    ''
    column)
```

However, constituents that might ordinarily be called NP present more of a problem, since noun modifiers are usually not annotated. The most common strategies are to leave the modifier flat (if simple) or to use NAC-TTL.

```
(NP The `` Thin Man '' series)


(NP the old Warner Bros. `` Road Runner '' cartoons)


(NP a
    (NAC-TTL `` My Favorite Bureaucrat '')
    plaque)


(NP the Oct. 20 ``
    (NAC-TTL Corporate Elite)
    '' issue)


(NP-SBJ The Wall Street Journal ``
        (NAC-TTL American Way
                 (PP of
                     (NP Buying)))
        '' Survey)


(NP a ``
    (NAC-TTL (NP Points)
             (PP of
                 (NP Light)))
    '' foundation)


(NP (NP the auto company 's)
    ``
    (NAC-TTL Cars
             (SBAR (WHNP-1 That)
                   (S (NP-SBJ *T*-1)
                      (VP Make
                          (NP Sense)))))
    ''
    campaign)
```

13 Gerunds and Participles

13.1 General remarks

13.1.1 Distributional distinction

It was decided that a theory-based distinction between nominal *-ing* clauses (gerunds) and other present participles would be too difficult to make consistently across annotators and throughout the corpus. Therefore, the distinction made in Treebank bracketing is a purely distributional one: *-ing* clauses are labeled S-NOM in subject position and as the object of a preposition, VP as the complement of *be*, S as the complement of other verbs, and S-ADV (or other appropriate adverbial tag) when modifying the matrix VP or sentence.

Since *-ing* clauses labeled S-NOM, S, and S-ADV/etc. are at least partly sentential in nature, like all other sentences they have subjects, either overt subjects or null * subjects when there is no overt subject present. They may also have VP-level complements and/or modifiers. The annotation of present and past participles is such that predicate-argument structure can be extracted from them as with ordinary sentences.

Past participles are labeled S (never S-NOM), with coindexing of the subject and adverbial function tags as appropriate.

(For the sake of convenience, the term "gerund" is used below to refer loosely to *-ing* clauses in general.)

13.1.2 Function tags

-ing clauses labeled S may receive the following tags: S-NOM-SBJ in subject position; S-NOM after prepositions; S after verbs and subordinating conjunctions; S-ADV (or -TMP, -LOC, -PRP, etc.) for adverbial functions.

13.1.3 Coindexation of null subjects

If there is no overt subject of the *-ing* clause, a null subject is present in the annotation: (NP-SBJ *). The null subject of an *-ing* clause is coindexed to another NP in the sentence if a coindexed interpretation is available. Coindexation proceeds as usual, according to pragmatic coreference as well as syntactic binding and control, and independent of the S-NOM/S distinction. However, null subjects of gerund complements of PP modifiers of NPs are coindexed only if there is a particularly strong coindexed interpretation. See section 4 [Null Elements] for more on the coindexation of null elements.

13.2 Present progressive

Any *-ing* form after auxiliary *be* is labeled VP and annotated as a complement. See section 13.3 for the annotation of present participles following other verbs.

```
(S (NP-SBJ I)
   (VP am
       (VP baking
           (NP cookies))))
```

13.3 Present participles

13.3.1 Subjects

Overt subjects, whether possessive or not, are bracketed as the subject of the *-ing* clause if the clause is labeled S. (If the clause is a gerund labeled S-NOM, the possessive is treated like any other possessive in NP.)

1. Subject of sentence

```
(S (SBAR-PRP Because
             (S (NP-SBJ-6 he)
                (VP should
                    (VP have
                        (VP been
                            (VP disqualified
                                (NP *-6)))))))
   ,
   (S-NOM-SBJ (NP-SBJ his)
              (VP playing
                  (ADVP at all)))
   (VP stinks))
```

2. *-ing* clause following verb

```
(S (NP-SBJ I)
   (VP do
       n't
       (VP mind
           (S (NP-SBJ you)
              (VP washing
                  (NP the car))))))
```

13.3.2 S-NOM vs. S

For the most part, there is not a theory-based distinction between nominal *-ing* clauses and other present participles, but rather a distributional one (see section 13.1.1 for the distributional distinction that is made). This section addresses the way in which S-NOM and S are used.

The S-NOM vs. S distinction for *-ing* clauses is made according to the following distributional criteria:

S-NOM.

-ing clauses are labeled S-NOM when they occur in the following positions:

1. as subjects of S (labeled S-NOM-SBJ)

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Baking
                  (NP pies)))
   (VP is
       (ADJP-PRD fun)))
```

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Walking
                  (ADVP-MNR quickly)))
   (VP is
       (NP-PRD good exercise)))
```

```
(S (SBAR-PRP Because
             (S (NP-SBJ-6 he)
                (VP should
                    (VP have
                        (VP been
                            (VP disqualified
                                (NP *-6)))))))
   ,
   (S-NOM-SBJ (NP-SBJ his)
              (VP playing
                  (ADVP at all)))
   (VP stinks))
```

2. as objects of prepositions (labeled S-NOM).

   Note that "preposition" here means any preposition that takes an NP on at least one of its uses. So *of, by, after, before, with, as, in,* etc. all head PPs when followed by an *-ing* clause, but *while, if* and other necessarily sentential subordinators are always SBAR. Note that all *-ing* clause objects of PP are S-NOM.

```
(S (NP-MNR That way)
   (NP-SBJ-7 investors)
   (VP can
       (ADVP essentially)
       (VP buy
           (NP the funds)
           (PP without
               (S-NOM (NP-SBJ *-7)
                      (VP paying
                          (NP the premium)))))))
```

   If an *-ing* clause has a null * subject, and there is a pragmatic coindexed interpretation, the subject is coindexed.

   Note that there is generally coindexation when the PP is an adjoined postmodifier of NP only if there is a particularly strong coindexed interpretation.

```
(S (NP-SBJ the company)
   (VP has
       (NP (NP no intention)
           (PP of
               (S-NOM (NP-SBJ *)
                      (VP tapping
                          (NP its short-term bank lines)))))
       (PP-TMP for
               (NP (NP a good part)
                   (PP of
                       (NP 1990))))))
```

Some examples:

- *about*

```
(NP (NP no squeamishness)
    (PP about
        (S-NOM (NP-SBJ *)
               (VP admitting
                   (NP this)))))
```

- *after*

```
(S (PP-TMP After
           (S-NOM (NP-SBJ *-1)
                  (VP winning
                      (NP the race))))
   ,
   (NP-SBJ-1 she)
   (VP ran
       (NP a victory lap)))
```

- *as*

```
(S (ADVP-TMP Often)
   (NP-SBJ the displeased parties)
   (VP interpreted
       (NP our decision)
       (PP-CLR as
               (S-NOM (NP-SBJ *)
                      (VP implying
                          (NP (NP favoritism)
                              (PP toward
                                  (NP the other))))))))
```

- *at*

```
(S (NP-SBJ-6 The government)
   (VP aimed
       (PP-CLR at
               (S-NOM (NP-SBJ *-6)
                      (VP stimulating
                          (NP (NP a faster rate)
                              (PP of
                                  (NP (NP economic growth)
                                      (PP of
                                          (NP the country))))))))))
```

- *before*

```
(S (PP-TMP Before
           (S (NP-SBJ *-1)
              (VP leaving
                  (PP-CLR for
                          (NP school)))))
   ,
   (NP-SBJ-1 *)
   (VP eat
```

```
          (NP a good breakfast)))
```

- *by*

```
(S (NP-SBJ-1 He)
   (VP inherited
       (NP a fortune)
       (PP-MNR by
               (S-NOM (NP-SBJ *-1)
                      (ADVP-MNR brutally)
                      (VP murdering
                          (NP his brother))))))
```

- *for*

```
(SINV (ADVP-LOC-PRD-TPC-1 Here)
      (VP would
          (VP be
              (ADVP-LOC-PRD *T*-1)))
      (NP-SBJ (NP a powerful force)
              (PP for
                  (S-NOM (NP-SBJ *)
                         (VP raising
                             (NP business activity))))))
```

- *from*

```
(S (NP-SBJ The police)
   (VP kept
       (NP-1 him)
       (PP-CLR from
               (S-NOM (NP-SBJ *-1)
                      (ADVP actually)
                      (VP collecting
                          (NP the money))))))
```

- *in*

```
(S (NP-SBJ This)
   (VP results
       (PP-CLR in
               (S-NOM (NP-SBJ-1 a separate record)
                      (VP being
                          (VP made
                              (NP *-1)))))))
```

- *of*

```
(S (NP-SBJ-6 I)
   (VP am
       (ADJP-PRD tired
                 (PP of
                     (S-NOM (NP-SBJ *-6)
                            (VP writing
                                (NP lists)))))))
```

- *since*

```
(S (NP-SBJ-1 I)
   (VP have
       (VP worked
           (NP several odd jobs)
           (PP-TMP since
                   (S-NOM (NP-SBJ *-1)
                          (VP leaving
                              (NP school)))))))
```

- *while*

```
(S (NP-SBJ-1 The committee)
   (VP continued
       (NP its meeting)
       (SBAR-TMP while
                 (S (NP-SBJ *-1)
                    (VP eating
                        (NP lunch))))))
```

- *with* (in absolute *with*-constructions)

```
(S (PP With
       (S-NOM (NP-SBJ interest rates)
              (VP rising)))
   ,
   (NP-SBJ the market)
   (VP is
       (VP moving
           (ADVP-MNR slowly))))
```

Note that *with* is bracketed as SBAR if it is not followed by an S-NOM (present participle).

```
(SBAR-ADV with
          (S (NP-SBJ-1 the new understudy)
             (VP hired
                 (NP *-1))))


(SBAR-ADV with
          (S (NP-SBJ his boyfriend)
             (ADJP-PRD abroad)))
```

3. as a child of the VP coordinated with other NP objects.
   In this case, the *-ing* clause is labeled S-NOM so that the bracket label at the level of coordination is NP rather than UCP (see section 13.5 on coordination below).

S.

-ing clauses are labeled S when they occur in the following positions:

1. as children of VP.
   Complements are labeled S, while adjuncts receive an appropriate tag: S-ADV, S-MNR, etc. See section 13.3.3.

```
(S (NP-SBJ I)
   (VP do
```

```
         n't
         (VP mind
             (S (NP-SBJ your)
                (VP washing
                    (NP the car))))))

(S (S (SBAR-ADV If
                (S (NP-SBJ it)
                   (VP promotes
                       (NP fashion)
                       (ADVP-MNR too much))))
      ,
      (NP-SBJ-1 the shop)
      (VP risks
          (S (NP-SBJ *-1)
             (VP alienating
                 (NP its old-line customers)))))
   ;
   (S (PP-MNR by
              (S-NOM (NP-SBJ *-2)
                     (VP emphasizing
                         ``
                         (NP value))))
      ,
      ''
      (NP-SBJ-2 it)
      (VP risks
          (S (NP-SBJ *-2)
             (VP watering
                 (PRT down)
                 (NP its high-minded mystique))))))

(S (NP-SBJ-1 Mrs. Ward)
   (VP took
       (PRT over)
       (PP-TMP in
               (NP 1986))
       ,
       (S-ADV (NP-SBJ *-1)
              (VP becoming
                  (NP (NP (NP the school 's)
                          seventh principal)
                      (PP-TMP in
                              (NP 15 years)))))))
```

Exception: if the gerund is a child of the VP but is coordinated with other NP objects, it is labeled S-NOM so that the bracket label at the level of coordination is NP rather than UCP (see section 13.5 on coordination below).

2. after subordinating conjunctions (labeled S, with no adverbial tag)

   Note that "subordinating conjunction" here means any subordinator that can never take an ordinary NP object (e.g., *while, when, if,* etc.) Subordinating conjunctions are never followed by S-NOM.

3. at S-level in preverbal position (labeled S-ADV/etc.).

13.3.3 S vs. S-ADV

The S/S-ADV, etc. distinction is made according to the following distributional criteria:

S.

An *-ing* clause is labeled S with no adverbial function tag if it is the complement of a verb or occurs in a "serial verb" construction.

All *-ing* complements of verbs other than *be* (e.g., *begin, come, continue, deny, get, go, justify, keep, like, permit, sit, stand, start, stop,* etc.) are bracketed in this way. As usual, the null subject is coindexed with the matrix subject if there is a coindexed interpretation.

```
(S (NP-SBJ-1 I)
   (VP like
       (S (NP-SBJ *-1)
          (VP helping
              (NP children)))))
```

```
(S (NP-SBJ-1 he)
   (VP (VP broke
           (PRT out)
           (NP the go codes))
       and
       (VP tried
           (S (NP-SBJ-2 *-1)
              (VP to
                  (VP start
                      (S (NP-SBJ *-2)
                         (VP transmitting
                             (NP one)))))))))
```

Note that this parallels the treatment of infinitival complements of some of these verbs:

```
(S (NP-SBJ-2 He)
   (VP tried
       (S (NP-SBJ-3 *-2)
          (VP to
              (VP start
                  (S (NP-SBJ *-3)
                     (VP to
                         (VP transmit
                             (NP one))))))))
```

S-ADV.

-ing clauses are given adverbial function tags if they behave as adverbial modifiers of the matrix VP or S.

1. The appropriate adverbial tag is used (instead of -ADV) if applicable: -TMP, -PRP, -LOC, -MNR.

2. -CLR is used in some rare cases, listed here:

   - *spend/waste time/money* X-*ing*

```
(S (NP-SBJ-1 Digital)
   (VP has
       (VP spent
           (NP (QP almost $ 1 billion) *U*)
           (S-CLR (NP-SBJ *-1)
                  (VP developing
                      (NP the new technology))))))
```

   - *have problems/difficulty/trouble* X-*ing*.
     The most recent annotation policy for this construction is represented here in (a), but some occurrences of this construction may be annotated according to an older policy, as given in (b). About half of the (a) analyses include the coindexing shown here.

```
(a)  (S (NP-SBJ-1 everybody)
        (VP will
            (VP have
                (NP (NP a difficult time)
                    (S-NOM (NP-SBJ *-1)
                           (VP reaching
                               (NP their profit objectives)))))))

(b)  (S (NP-SBJ-1 everybody)
        (VP will
            (VP have
                (NP a difficult time)
                (S-CLR (NP-SBJ *-1)
                       (VP reaching
                           (NP their profit objectives))))))
```

3. -ADV is used if no other adverbial tag applies.
   This is especially common with floating participles, including dangling participles (see section 13.6.2 for more details on floating participles).

```
(S (S-ADV (NP-SBJ God)
          (VP willing))
   ,
   (NP-SBJ we)
   (VP will
       (VP arrive
           (PP-LOC at
                   (NP our destination))
           (ADVP-MNR safely))))

(S (S-ADV (NP-SBJ *-1)
          (VP Having
              (ADVP-MNR carefully)
              (VP considered
                  (NP his options))))
   ,
   (NP-SBJ-1 he)
   (VP decided
```

```
(S (NP-SBJ *-1)
   (VP to
       (VP take
           (NP the job))))))
```

4. All past participles that modify S or VP are labeled S-ADV/etc.

```
(S (S-ADV (NP-SBJ-1 *-2)
          (VP Given
              (NP *-1)
              (NP the chance)))
   ,
   (NP-SBJ-2 I)
   (VP 'd
       (VP do
           (NP it)
           (ADVP-TMP again))))
```

5. Dangling participles. "Floating participles" here includes dangling participles. They are labeled S-ADV and a null * subject is coindexed as appropriate.

```
(S (S-ADV (NP-SBJ *-1)
          (VP Living
              (PP-LOC-CLR in
                          (NP this house))))
   ,
   (NP-SBJ the noise)
   (VP is
       (VP driving
           (S (NP-SBJ-1 me)
              (ADJP-PRD buggy))))
   !)
```

13.3.4 NP vs. S or S-NOM

Single-word nominal *-ing* clauses are labeled NP; the only exception is for those with a strong event reading. Gerunds that have an overt subject or a complement or are modified by an adverbial are bracketed as VP dominated by S or S-NOM.

1. Distinguishing between NP and S or S-NOM

   A distinction is made between nouns ending in *-ing* (labeled NP, with the head generally POS-tagged NN) and *-ing* clauses (labeled S or S-NOM, with the head generally POS-tagged VBG), according to the following criteria:

   (a) An *-ing* form is labeled NP if it:

      i. is a single word (e.g., *running*), except when it has a strong event reading (see page 221).
      ii. has a determiner (e.g., *some teaching*)
      iii. has an *of* PP object (e.g., *teaching of difficult students*)
      iv. has other modifiers that could be modifying an ordinary noun (e.g., *world-class running*), as opposed to adverbial modifiers, which suggest VP.

      - Subject position:

```
(S (NP-SBJ Baking)
   (VP is
       (ADJP-PRD fun)))
```

- Object position:

```
(S (NP-SBJ I)
   (VP like
       (NP (NP field hockey)
           and
           (NP swimming))))
```

- With an NP possessive:

```
(S (NP-SBJ The men)
   (VP were
       (ADJP-PRD tired)
       (PP from
           (NP (NP a night 's)
               drinking))))
```

- With a possessive pronoun:

```
(S (NP-SBJ We)
   (VP kicked
       (NP him)
       (PP-CLR out
               (PP of
                   (NP the band)))
       (SBAR-PRP because
                 (S (NP-SBJ his playing)
                    (VP stinks)))))
```

- With quantifiers:

```
(S (NP-SBJ There)
   (VP is
       (NP-PRD no smoking)
       (PP-LOC on
               (NP this flight))))

(S (NP-SBJ There)
   (VP will
       (VP be
           (NP-PRD no talking)
           (PP-TMP during
                   (NP the movie)))))
```

- With PP postmodifiers:

```
(S (NP-SBJ There)
   (VP 's
       (VP been
           (NP-PRD (NP no finding)
                   (PP by
                       (NP anybody))
```

```
                    (PP of
                        (NP (NP any substantive violation)
                            (PP of
                                (NP any antitrust laws))))))))
```

Gerunds with *of* complements and other adjectival and PP modifiers of NP are bracketed just like ordinary NPs:

```
(S (NP-SBJ (NP The taking)
           (PP of
               (NP Iwo Jima)))
   (VP was
       (NP-PRD no easy feat)))
```

contrast with:

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Taking
                  (NP Iwo Jima)))
   (VP was
       (NP-PRD no easy feat)))
```

- With a non-PP postmodifier:

```
(S (NP-SBJ (NP The dancing)
           ,
           (SBAR (WHNP-7 which)
                 (S (NP-SBJ *T*-7)
                    (VP was
                        (ADJP-PRD very good))))
           ,)
   (VP began
       (PP-TMP at
               (NP 8:00))))
```

(b) An *-ing* form with a strong event reading may be labeled S or S-NOM:
Single-word gerund objects of verbs are normally labeled NP, but in sentences with strong event readings they may be labeled S or S-NOM. The null subjects of these gerunds are coindexed if appropriate.

```
(S (NP-SBJ Apple)
   (ADVP allegedly)
   (VP discouraged
       (NP-2 retailers)
       (PP-CLR from
               (S-NOM (NP-SBJ *-2)
                      (VP discounting)))))
```

The DEFAULT is to label as NP, as in (a) below. For example, the expression *I hate lying* has both an NP ("I hate it when others lie") interpretation, as in (a), and an S interpretation ("I hate to lie"), as in (b):

(a) "I hate it when others lie"

```
(S (NP-SBJ I)
   (VP hate
```

```
(NP lying)))
```

(b) "I hate to lie"

```
(S (NP-SBJ-1 I)
   (VP hate
       (S (NP-SBJ *-1)
          (VP lying))))
```

2. Overt subjects

   When the gerund has an overt genitive subject, as in *They liked our singing*, it is labeled NP unless it clearly warrants a clausal interpretation.

   (a) "They liked the way we sang"

```
(S (NP-SBJ They)
   (VP liked
       (NP our singing)))
```

   (b) "They liked the fact that we sang"

```
(S (NP-SBJ They)
   (VP liked
       (S (NP-SBJ our)
          (VP singing))))
```

   (c) However, if the subject is not possessive, it is always bracketed S:

```
(S (NP-SBJ we)
   (VP have
       (S (NP-SBJ them)
          (VP waiting))))
```

3. Additional complements, modifiers, etc.

   In cases where the gerund occurs with complements, modifiers, etc., S or S-NOM is used only when the structure is unambiguously clausal because the gerund has a direct object. Otherwise, NP is used.

```
(S (NP-SBJ They)
   (VP liked
       (S (NP-SBJ our)
          (VP singing
              (NP folk songs)))))


(S (S-NOM-SBJ (NP-SBJ *)
              (VP Taking
                  (NP Iwo Jima)))
   (VP was
       (NP-PRD no easy feat)))
```

4. Quantifiers.

   Any quantifier can be a determiner for an S-NOM *-ing* clause. If the gerund is labeled S-NOM, the S-NOM and the (unlabeled) quantifier are both children of an outer NP label.

```
(S (NP-SBJ There)
   (VP was
       (ADVP certainly)
       (NP-PRD no
               (S-NOM (NP-SBJ *)
                      (VP stopping
                          (NP (NP the tide)
                              (PP of
                                  (NP emotion))))))))
```

13.3.5 ADJP vs. S

A gerund may be ambiguous between two interpretations: adjectival noun phrase modifier (ADJP), and gerund clause (S). For instance, the sentence *Flying planes can be dangerous* has at least two interpretations in isolation, paraphrased below. When it is not clear whether a gerund should be analyzed as ADJP or S, the DEFAULT is to analyze it as an adjective, as in (a).

(a) "Planes which are flying (overhead) can be dangerous"

```
(S (NP-SBJ Flying planes)
   (VP can
       (VP be
           (ADJP-PRD dangerous))))
```

(b) "The act of flying planes can be dangerous"

```
(S (S-NOM-SBJ (NP-SBJ *)
              (VP Flying
                  (NP planes)))
   (VP can
       (VP be
           (ADJP-PRD dangerous))))
```

13.3.6 Pseudo-prepositions

"Pseudo-prepositions" are words that behave like prepositions but are historically or apparently verb participles.

Tests for deciding whether a gerund is a "pseudo-preposition":

1. Pied-piping.

   Members of the class of "pseudo-prepositions" are admitted if they can undergo pied piping, and only if the annotator has a strong intuition that the item in question is behaving as a preposition.

2. Lack of verbal content.

   These pseudo-prepositions either have no real verbal meaning or have a meaning other than their ordinary verbal usage.

3. POS tagging.

   While these pseudo-prepositions are **bracketed** with a PP label, the Part of Speech tags associated with these words are still VBG (gerund verb) or VBN (past participle verb), as described in the *POS guidelines* [Santorini 1990].

```
(PP (VBG including)
    (NP (DT the)
        (NN kitchen)
        (NN sink)))
```

4. Real prepositions. Note that while *during* and *pending* may at first glance look like participles, they lack corresponding verbs, so these should be POS-tagged IN or JJ, according to usage. However, automatic tagging tools tend to assign a VBG tag, so these may occasionally be erroneously analyzed as verbs.

5. Examples. The following is a partial list of items annotated as pseudo-prepositions:

   according to, barring, based on, combined with, compared with, concerning, depending on, excluding, following, given, including, provided (that), regarding

   - *according to*

   ```
   (S (PP According
          (PP to
              (NP (NP sources)
                  (PP-LOC in
                          (NP the White House)))))
      ,
      (NP-SBJ the President)
      (VP has
          (VP been
              (ADJP-PRD depressed)
              (PP-TMP since
                      (NP the election)))))
   ```

   Test: According to whom has the President been depressed?

   - *barring*

   ```
   (S (PP barring
          (NP (NP a recession)
              and
              (NP (NP a further strengthening)
                  (PP of
                      (NP the dollar))
                  (PP against
                      (NP foreign currencies)))))
      ,
      (NP-SBJ the industry)
      (VP is n't
          (VP headed
              (PP-CLR for
                      (NP a prolonged slump)))))
   ```

   - *based on*

   ```
   (VP staggering
       (NP rates)
       (PP based
           (PP on
               (NP (NP the size)
   ```

```
                  (PP of
                      (NP deposit))))))
```

- *combined with*

```
(S (NP-SBJ (NP The
               (NX (NX glut)
                   and
                   (NX consequent lower prices)))
           ,
           (PP combined
               (PP with
                   (NP cancer fears)))
           ,)
   (VP was
       (NP-PRD (NP a (ADJP very serious) blow)
               (PP to
                   (NP growers)))))
```

- *compared with*

```
(S (NP-SBJ IBM stock)
   (VP sold
       (PP-CLR at
               (NP $1.25))
       ,
       (PP compared
           (PP with
               (NP (NP $1.32)
                   (ADVP-TMP (NP a month)
                             ago))))))
```

Test: Compared with what did IBM stock sell at $1.25? Note also that this use of *compared with* differs from its verbal meaning.

- *concerning* (when equivalent to *about*)

```
(S (NP-SBJ Imogen)
   (VP admitted
       (NP (NP a mild curiosity)
           (PP concerning
               (NP Flavia)))))
```

Test: Concerning whom did Imogen admit a mild curiosity?

- *depending on*

```
(S (PP Depending
       (PP on
           (NP the organism)))
   ,
   (NP-SBJ there)
   (VP may
       (VP be
           (NP-PRD multiplication)
           (PP-LOC in
                   (NP some food or beverage products)))))
```

Contrast with its verbal use:

```
(S (S-ADV (NP-SBJ *-1)
          (VP Depending
              (PP-CLR on
                      (NP (NP the babysitter 's)
                          reliability))))
   ,
   (NP-SBJ-1 they)
   (VP stayed
       (ADVP-LOC-CLR out)
       (ADVP-TMP late)))
```

- *excluding*

```
(S (NP-SBJ (NP net sales)
           (PP of
               (NP (NP all mutual funds)
                   ,
                   (PP excluding
                       (NP money market funds))
                   ,)))
   (VP fell
       (PP-DIR to
               (NP (QP $ 1.9 billion) *U*)
               (PP-TMP in
                       (NP September)))
       (PP-DIR from
               (NP (QP $ 4.2 billion) *U*)
               (PP-TMP in
                       (NP August)))))
```

- *following* (when equivalent to *after*)

```
(S (NP-SBJ Soviet police)
   (VP clashed
       (PP-CLR with
               (NP demonstrators))
       (PP-LOC in
               (NP Moscow))
       (PP-TMP following
               (NP (NP a candlelight vigil)
                   (PP-LOC around
                           (NP (NP the KGB 's)
                               Lubyanka headquarters))))))
```

Test: Following what did the Soviet police clash with demonstrators?

Contrast with its verbal use:

```
(S (S-ADV (NP-SBJ *-1)
          (VP Following
              (NP (NP the doctor 's)
                  directions)))
   ,
   (NP-SBJ-1 she)
```

```
(VP took
    (NP one pill)
    (PP-TMP after
            (NP each meal))))
```

- *given* (when it means *in light of* or *considering*)

```
(S (PP Given
       (NP the present conditions))
   ,
   (NP-SBJ I)
   (VP think
       (SBAR 0
             (S (NP-SBJ she)
                (VP 's
                    (VP done
                        (ADVP-MNR rather well)))))))
```

Contrast with verbal use:

```
(S (S-ADV (NP-SBJ-1 *-2)
          (VP Given
              (NP *-1)
              (NP the chance)))
   ,
   (NP-SBJ-2 I)
   (VP 'd
       (VP do
           (NP it)
           (ADVP-TMP again))))
```

- *including*

```
(PP including
    (NP the kitchen sink))
```

- *provided (that)*

Note that although the pied piping part of the pseudo-prepostionhood test doesn't work with *that*-clauses, the other criterion ("lack of verbal meaning") is applicable. In cases where the *that* is absent but interpreted, SBAR 0 is inserted:

```
(S (NP-SBJ The prepaid plans)
   (VP may
       (VP be
           (NP-PRD a good bet)
           ,
           (PP provided
               (SBAR 0
                     (S (NP-SBJ (NP the guarantee)
                                (PP of
                                    (NP future tuition)))
                        (VP is
                            (ADJP-PRD secure)))))))
```

- *regarding* (when it means *about*)

```
(S (NP-SBJ I)
   (VP need
       (NP (NP some information)
           (PP regarding
               (NP (NP flights)
                   (PP-DIR to
                           (NP Guam)))))))
```

6. Multi-word prepositions.

   The above pseudo-prepositions should not be confused with "multi-word" prepositions, which are bracketed flat. The following is an exhaustive list of multi-word prepositions: *because of*, *instead of*, *rather than*, and *such as*. (See section section 26 [Orphans].)

13.4 Past Participles

Past participles are labeled S, and adverbial function tags are added if appropriate.

13.4.1 Prepositions

A preposition or subordinator that dominates a past participial clause is bracketed as SBAR. The rule about prepositions (as described in section 2 on page 212) applies only to *-ing* clause complements, not to other sentential complements (i.e., *until* is always a PP with gerund complements because it **could** take an NP complement: *until yesterday/last year*/etc. However, since past participles are always Ss, *until* with a past participle complement is an SBAR).

13.4.2 Function tags

If not under a subordinator, the participle receives the appropriate function tag.

```
(S (S-ADV (NP-SBJ-1 *-2)
          (VP Given
              (NP *-1)
              (NP the chance)))
   ,
   (NP-SBJ-2 I)
   (VP 'd
       (VP do
           (NP it)
           (ADVP-TMP again))))
```

If under a subordinator (here, *until*), it is the subordinator that bears the adverbial function tag.

```
(S (NP-SBJ-1 I)
   (VP will
       (VP wait
           (ADVP-LOC here)
           (SBAR-TMP until
                     (S (NP-SBJ-2 *-1)
                        (VP asked
```

```
(NP-3 *-2)
(S (NP-SBJ *-3)
   (VP to
       (VP leave))))))))
```

13.4.3 Coindexation and tracing

Coindexation and tracing proceed as usual with past participles. This means that there is generally a passive trace coindexed with a null subject in participial clauses. The null subject is coindexed with another NP in the sentence if appropriate, according to interpretation.

13.5 Coordination

Coordination rules laid out in section 7 [Coordination] apply to NP, S-NOM, S, S-ADV with no change. See section 8 [Shared Complements and Modifiers] for more details on the annotation of coordinate structures and shared elements.

There are only two special cases in the coordination of gerunds and participles — the coordination of S-NOM with NP, and S becoming S-NOM in order to coordinate with NP.

1. Level of coordination is labeled NP.

   The coordination of S-NOM and NP is labeled NP.

```
(S (NP-SBJ we)
   (VP must
       (VP choose
           (PP-CLR between
                   (NP (NP peace)
                       and
                       (S-NOM (NP-SBJ *)
                              (VP keeping
                                  (NP the Communists)
                                  (PP-LOC-CLR out
                                              (PP of
                                                  (NP Berlin)
                                                  )))))))))
```

2. S becomes S-NOM.

   Although the *-ing* clause object in VP is normally labeled S, it is labeled S-NOM when it is coordinated with another NP object so that it can be coordinated under an NP label. Coindexation of the null * subject is less likely than usual.

```
(S (NP-SBJ I)
   (VP like
       (NP (NP cookies)
           ,
           (NP mako sharks)
           ,
           and
           (S-NOM (NP-SBJ *)
                  (VP swimming
```

```
                    (PP-LOC in
                            (NP the lake))
                    (PP-TMP on
                            (NP Tuesdays)))))))
```

If it is necessary to coordinate any S other than a gerund with an NP, the coordination must be labeled UCP (Unlike Coordinated Phrase).

```
(S (NP-SBJ-5 she)
   (VP gets
       (UCP (NP the best comic bits)
            and
            (S (NP-SBJ *-5)
               (VP to
                   (VP wear
                       (NP glamorous dresses)))))))
```

13.6 Reduced relatives and floating participles

There are two kinds of NP-modifying participle — the "reduced relative" and the "floating participle". Reduced relatives are those that are closely related to the NP, and don't easily appear in any position other than just after the modified NP. This type tends to roughly correspond to **restrictive** relatives (but note that even this rough distinction is **not** made for non-reduced relatives). Floating participles are those that move easily around the sentence (beginning and end of sentence, as well as just after the noun).

13.6.1 Reduced Relative Clause

The reduced relative clause resembles a restrictive relative clause in which the complementizer and auxiliary verb are absent. It postmodifies a noun and consists of either a past participle, a present participle, or an ADJP, NP, or PP with sentential modifiers.

VP. If the reduced relative is a past or present participle, the participle is labeled VP and adjoined to the NP.

```
(S (NP-SBJ-2 (NP An orangutan)
             (VP foaming
                 (PP-CLR at
                         (NP the mouth))))
   (VP should
       not
       (VP be
           (VP provoked
               (NP *-2)))))
```

In the case of passives, the passive trace is indicated by (NP *). However, note that this null element does not bear an index, as ordinarily it would be coindexed with the subject of the relative clause, which in this case is not present in the annotation.

```
(S (NP-SBJ He)
   (VP bought
       (NP (NP two watches)
```

```
          (VP designed
              (NP *)
              (PP by
                  (NP-LGS Paloma Picasso))))))
```

Notice that the passive trace (NP *) may sometimes function as the subject of a subordinate clause:

```
(NP (NP an elephant)
    (VP called
        (S (NP-SBJ *)
           (NP-PRD Dumbo))
        (SBAR-PRD because
                  (S (NP-SBJ his ears)
                     (VP were
                         (ADJP-PRD so large))))))
```

Non-VP: RRC and other likely common alternates. The label RRC is used only if the "reduced relative" is not a VP, but rather some other postmodifier such as NP, PP, ADJP, or ADVP that itself has "sentential" modifiers. The RRC bracketing provides an additional level under which to attach these modifiers. (Note that use of RRC is rare.)

```
(NP (NP 110 titles)
    (RRC not
         (ADVP-TMP presently)
         (PP-LOC in
                 (NP the collection))))


(NP (NP the negative ad)
    ,
    (RRC (PP-TMP for
                 (NP years))
         (NP (NP a secondary presence)
             (PP-LOC in
                     (NP most political campaigns))))
    ,)


(NP (NP this kind)
    (PP of
        (NP mudslinging))
    ,
    (RRC (ADVP-TMP long)
         (ADJP empty
               (PP of
                   (NP significant issues))))
    ,)
```

However, note that despite this policy and despite the fact that reduced relatives of this type are fairly widespread in the corpus, the RRC label is in general **not** used by the annotators. Instead, one of several alternate annotations may be found, illustrated here.

1. where modifiers are bracketed as children of the modifying phrase.

```
(NP (NP 110 titles)
    (PP-LOC not
            (ADVP-TMP presently)
            in
            (NP the collection)))


(NP (NP the negative ad)
    ,
    (NP (NP (ADVP-TMP always)
            a secondary presence)
        (PP-LOC in
                (NP most political campaigns)))
    ,)


(NP (NP this kind)
    (PP of
        (NP mudslinging))
    ,
    (ADJP (ADVP-TMP long)
          empty
          (PP of
              (NP significant issues)))
    ,)
```

2. where modifiers are adjoined to the modifying phrase.

```
(NP (NP the negative ad)
    ,
    (NP (PP-TMP for
                (NP years))
        (NP a secondary presence)
        (PP-LOC in
                (NP most political campaigns)))
    ,)


(NP (NP this kind)
    (PP of
        (NP mudslinging))
    ,
    (ADJP (ADVP-TMP long)
          (ADJP empty
                (PP of
                    (NP significant issues))))
    ,)
```

3. where modifiers are adjoined separately to the NP (as though they were modifiers of the NP rather than modifiers of the modifying phrase itself). (This option, which in fact misrepresents the semantic structure, is rare.)

```
(NP (NP the books)
    (PP-LOC on
            (NP the shelf))
    (NP-TMP yesterday))


(NP (NP the negative ad)
    ,
    (PP-TMP for
            (NP years))
    (NP (NP a secondary presence)
        (PP-LOC in
                (NP most political campaigns)))
    ,)


(NP (NP this kind)
    (PP of
        (NP mudslinging))
    ,
    (ADVP-TMP long)
    (ADJP empty
          (PP of
              (NP significant issues)))
    ,)
```

4. which represents the **coordination** of reduced relatives, with modifiers placed at the level of coordination.

```
(NP (NP this kind)
    (PP of
        (NP mudslinging))
    ,
    (ADJP (ADVP-TMP long)
          (ADJP empty
                (PP of
                    (NP significant issues)))
          ,
          but
          (ADVP-TMP still)
          (ADJP common)))
```

13.6.2 Floating participles

"Floating participle" is a blanket term used by the Treebank to refer to a modifying predicate attached at S or VP level. They include past participles, present participles/gerunds, adjectives, and the occasional NP or PP. They are bracketed as VPs or -PRDs dominated by an S-ADV.

Floating participles are placed at S-level if they occur before the verb and at VP-level if they occur after the verb. They are labeled S-ADV, often with a null subject that is coindexed with the appropriate NP (usually the subject of the matrix clause).

1. Before the subject.

   When the floating participle appears before the subject, it is never analyzed as a reduced relative and is bracketed as follows:

```
(S (S-ADV (NP-SBJ *-1)
          (ADJP-PRD heady
                    (PP with
                        (NP success))))
   ,
   (NP-SBJ-1 I)
   (VP rushed
       (NP it)
       (PRT in))))
```

```
( (S-3 (S-ADV (NP-SBJ-2 *-1)
              (VP Considered
                  (NP *-2)
                  (PP-CLR as
                          (NP a whole))))
       (PRN ,
            (S (NP-SBJ Mr. Lane)
               (VP said
                   (SBAR 0
                         (S *T*-3))))
            ,)
       (NP-SBJ-1 the filings)
       ``
       (VP will
           (VP be
               (ADJP-PRD effective)))
       . ''))
```

```
( (S (S-ADV (NP-SBJ-2 *-1)
            (VP Clad
                (NP *-2)
                (PP-CLR in
                        (NP his trademark black velvet suit))))
     ,
     (NP-SBJ-1 the soft-spoken clarinetist)
     (VP announced
         (SBAR that
               (S (NP-SBJ-9 (NP his new album)
                            , ``
                            (NP-TTL Inner Voices)
                            , '')
                  (VP had
                      (ADVP-TMP just)
                      (VP been
                          (VP released
                              (NP *-9)))))))
     .))
```

2. After the subject; nonadjacent to the subject.

   When the floating participle appears after the subject and nonadjacent to the subject, it is never analyzed as a reduced relative:

```
(S (NP-SBJ-1 I)
   (VP rushed
       (NP it)
       (PRT in)
       ,
       (S-ADV (NP-SBJ *-1)
              (ADJP-PRD heady
                        (PP with
                            (NP success))))))
```

3. After the subject; adjacent to the subject.

   If the participle appears after the subject but adjacent to it or its modifiers, it can be bracketed either as a reduced relative or as a floating participle. Annotators use the following tests to decide whether a given modifier is a reduced relative or a floating participle. (Of course, which test(s) the annotator decides to use will influence the eventual annotation. In most cases, the outcome is the same, but there is some variation in the results.) The tests are listed in the order that they are most likely to be used by the annotators.

   Distinguishing floating participle from reduced relative:

   (a) Mobility.

   Is its semantic relation to the sentence maintained if it is moved around the sentence?
   Yes → floating participle
   No → reduced relative
   "??Reported by the advisory committee, the progress is real"
   This participial clause cannot move; therefore, reduced relative.

```
(S (NP-SBJ (NP The progress)
           (VP reported
               (NP *)
               (PP by
                   (NP-LGS the advisory committee))))
   (VP is
       (ADJP-PRD real)))
```

   (b) Commas.

   Does it have comma intonation (insofar as that can be ascertained)? There is a strong tendency for the choice to be influenced by the presence of commas in the text, which signal comma intonation in the case of floating participles.
   Yes → floating participle
   No → reduced relative
   "And now, the woman, tired and trembling, came here..."
   This participial clause requires comma intonation; therefore, floating participle.

```
(S And
   (ADVP-TMP now)
   ,
   (NP-SBJ-1 the woman)
   ,
```

```
(S-ADV (NP-SBJ *-1)
       (UCP-PRD (ADJP-PRD tired)
                and
                (VP trembling)))
,
(VP came
    (ADVP-DIR here)
    (PP-DIR to
            (NP the DeKalb County cannery))))
```

(c) Paraphrase with *while* or *being*.
Is its semantic relation to the sentence maintained if *while* or *being* is inserted?
Yes → floating participle
No → reduced relative
"The Rusk belief, while replacing..."
This is OK; therefore, floating participle.

```
(S (NP-SBJ-1 (NP The Rusk belief)
             (PP in
                 (NP balanced defense)))
   ,
   (S-ADV (NP-SBJ *-1)
          (VP replacing
              (NP (NP the Dulles theory)
                  (PP of
                      (NP massive retaliation)))))
   ,
   (VP removes
       (NP a grave danger)))
```

"The progress, while reported..." This is not OK; therefore, reduced relative.

```
(S (NP-SBJ (NP The progress)
           (VP reported
               (NP *)
               (PP by
                   (NP-LGS the advisory committee))))
   (VP is
       (ADJP-PRD real)))
```

(d) Paraphrase with *which/that is/are*.
Is its semantic relation to the sentence maintained if *which/that is/are* is inserted?
Yes → reduced relative
No → floating participle
"The progress that is reported..."
This is good; therefore, reduced relative.

```
(S (NP-SBJ (NP The progress)
           (VP reported
               (NP *)
               (PP by
                   (NP-LGS the advisory committee))))
   (VP is
```

```
            (ADJP-PRD real)))
```

4. Non-participial S-ADVs.

   Note that although the vast majority of S-ADVs are in fact "floating participles", any adverbial modifier that contains a predicate merits S + the relevant adverbial tag(s), for example infinitival clauses and things such as *I hope not too late*:

```
(S (S (NP-SBJ She)
      (VP asked
          (SBAR if
                (S (NP-SBJ I)
                   (VP had
                       (NP other advice))))))
   and
   ,
   (S (NP-SBJ I)
      (VP rushed
          (NP it)
          (PRT in)
          ,
          (S-ADV (NP-SBJ I)
                 (VP hope
                     (SBAR 0
                           (S (NP-SBJ *)
                              (ADJP-PRD not too late))))))))
```

14 Infinitives

14.1 Bare infinitives

14.1.1 Complements of verbs

Bare infinitives (i.e. ones without *to* preceding them) are bracketed as VPs. The bare infinitive complements of perception verbs (*see, hear, feel*) and causative verbs (*make, let, have*; also *help*) are bracketed together with the NP preceding them as a complement S. The structural subjects of both the matrix clause and the embedded clause are tagged -SBJ.

```
(S (NP-SBJ I)
   (VP saw
       (S (NP-SBJ him)
          (VP do
              (NP it)))))
```

14.1.2 Imperatives

Imperatives are bracketed similarly. The addressee, if present in the sentence, is tagged -VOC. Note that the * subject is not coindexed with the -VOC phrase.

```
(S (NP-VOC Junior)
   ,
   (NP-SBJ *)
   (VP finish
       (NP your dinner)))
```

14.2 *To* infinitives

These fall into quite a few categories, but the basic structure of an infinitival with *to* is:

```
(S (NP-SBJ <subject>)
   (VP to
       (VP <main verb> ...)))
```

The <subject> may be either overt, a null *, or in certain relative clauses, a null *T*.

14.2.1 Infinitival relative clauses (IRCs)

The null * subject of an IRC is never indexed, except when the gap is in subject position, in which case the WHNP element is coindexed with a *T*.

1. Trace in verbal object position.

```
(NP (NP much research)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
```

```
                (VP do
                    (NP *T*-1))))))
```

2. Trace in prepositional object position.

```
(NP (NP a crisis)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP get
                     (PP-CLR past
                             (NP *T*-1))
                     (PP-TMP within (NP a few days)))))))
```

3. Trace in subject position.

- Active

```
(NP (NP bloodhounds)
    (SBAR (WHNP-4 0)
          (S (NP-SBJ *T*-4)
             (VP to
                 (VP trail
                     (NP the assassins))))))
```

- Passive

```
(NP (NP work)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ-2 *T*-1)
             (VP to
                 (VP be
                     (VP done
                         (NP *-2)))))))
```

4. With *for*.

```
(NP (NP a desirable organization)
    (SBAR (WHNP-3 0)
          for
          (S (NP-SBJ us)
             (VP to
                 (VP join
                     (NP *T*-3))))))
```

5. Trace in adjunct position.

```
(NP (NP time)
    (SBAR (WHADVP-1 0)
          (S (NP-SBJ *)
             (VP to
```

```
                (VP go
                    (ADVP-TMP *T*-1))))))
```

6. With *wh*-word.

```
(NP (NP three levels)
    (SBAR (WHPP-1 on
                  (WHNP which))
          (S (NP-SBJ *)
             (VP to
                 (VP treat
                     (NP the subject)
                     (PP-LOC *T*-1))))))
```

7. Some tricky cases.

   Despite the occasional intuition that the following infinitivals and others like them are complements, they are bracketed as adjunct relative clauses.

   - *the first...to...*

```
(S (NP-SBJ It)
   (VP was
       (NP-PRD (NP the first)
               (PP-LOC in
                       (NP the series))
               (SBAR (WHNP-1 0)
                     (S (NP-SBJ-2 *T*-1)
                        (VP to
                            (VP be
                                (VP sponsored
                                    (NP *-2)
                                    (PP by
                                        (NP-LGS First Lady
                                                Jacqueline Kennedy))))))))))
```

   - *the only...to...*

```
(S (NP-SBJ The National Gallery)
   (VP is
       (NP (NP the only museum)
           (PP-LOC in
                   (NP the country))
           (SBAR (WHNP-1 0)
                 (S (NP-SBJ *T*-1)
                    (VP to
                        (VP have
                            (NP a full-time music director))))))))
```

14.2.2 Complements of nouns

The infinitival complements of nouns do not get SBAR structure and are instead labeled S. The null subject in this case does not get an index. (See section 11 [Modification of NP] for more information about the bracketing of complements of nouns.)

```
(NP a brave attempt
    (S (NP-SBJ *)
       (VP to
           (VP describe
               (NP the English language)))))
```

14.2.3 Complements of adjectives/adverbs

The *likely* type (with null subject). Certain ("subject-raising") adjectives, such as *likely*, take an infinitival complement whose subject is interpreted as coreferent with the subject of the matrix clause. For these adjectives, the null * subject of the infinitive is coindexed to the matrix subject.

```
(S (NP-SBJ-4 This distinction)
   (VP is
       (ADJP-PRD likely
                 (S (NP-SBJ *-4)
                    (VP to
                        (VP be
                            (ADJP-PRD difficult)))))))
```

```
(S (NP-SBJ-1 Zaphod)
   (VP is
       (ADJP-PRD ready , willing , and able
                 (S (NP-SBJ *-1)
                    (VP to
                        (VP eat
                            (NP the steak)))))))
```

The *tough* type (with null object). Other adjectives, such as *tough*, take an infinitival complement that contains a gapped object (of PP or VP), which is represented with *T*, coindexed to a (WHNP 0) in an SBAR introducing the clause.

```
(S (NP-SBJ A pro)
   (VP is
       (ADJP-PRD tough/easy/hard
                 (SBAR (WHNP-1 0)
                       (S (NP-SBJ *)
                          (VP to
                              (VP beat
                                  (NP *T*-1)))))))))
```

```
(S (NP-SBJ (NP Sentences)
           (PP with ‘ (NP tough) ’))
   (VP are
```

```
        (ADJP-PRD tough
                  (SBAR (WHNP-1 0)
                        for
                        (S (NP-SBJ syntacticians)
                           (VP to
                               (VP explain
                                   (NP *T*-1))))))))
```

```
(S ``
   (NP-TTL-SBJ Psyche)
   ''
   ,
   (VP is
       (NP-PRD (NP a lush , sweet-sounding affair)
               (SBAR
                (WHNP-2 that)
                (S (NP-SBJ *T*-2)
                   (VP was
                       (ADJP-PRD pleasant
                                 (SBAR (WHNP-1 0)
                                       (S (NP-SBJ *)
                                          (VP to
                                              (VP encounter
                                                  (NP *T*-1)
                                                  (ADVP-TMP once
                                                            again))))))))))))
```

After *enough*. *Enough* introduces an infinitival clause with either a null or overt subject, and no gap otherwise.

```
(S (NP-SBJ There)
   (VP has
       not
       (VP been
           (NP-PRD (NP time)
                   (ADJP enough
                         (S (NP-SBJ *)
                            (VP to
                                (VP institute
                                    (NP reforms))))))))))
```

```
(S (NP-SBJ We)
   (VP want
       (S (NP-SBJ our policy)
          (VP to
              (VP be
                  (ADJP-PRD consistent enough
                            (S (NP-SBJ *)
                               (VP to
                                   (VP please
                                       (NP everyone))))))))))
```

```
(S (NP-SBJ-1 We)
   (VP need
       (S (NP-SBJ *-1)
          (VP to
              (VP put
                  (PRT in)
                  (NP enough gas
                      (S (NP-SBJ *)
                         (VP to
                             (VP reach
                                 (NP Reno))))))))))
```

```
(S (NP-SBJ-1 We)
   (VP need
       (S (NP-SBJ *-1)
          (VP to
              (VP put
                  (PRT in)
                  (NP enough gas
                      (SBAR for
                            (S (NP-SBJ the car)
                               (VP to
                                   (VP reach
                                       (NP Reno)))))))))))
```

After *too***.** *Too* in an adjective or adverb phrase often introduces an infinitival clause that indicates extent. Some of these have obvious gaps, and in others it's difficult to find one. If a gap is apparent, the infinitival is labeled SBAR, and a *T* in the position of the gap is coindexed with (WHNP 0) in SBAR, as in (a) and (b). If no gap is apparent, the infinitival is labeled S and attached under ADJP, as in (c).

- Object gap.

```
(S (NP-SBJ (NP The desolation)
           (PP of
               (NP a post-attack world)))
   (VP would
       (VP be
           (ADJP-PRD too awful
                     (SBAR (WHNP-1 0)
                           (S (NP-SBJ *)
                              (VP to
                                  (VP face
                                      (NP *T*-1)))))))))
```

- Subject gap.

```
(S (NP-SBJ (NP (NP Col. Faget 's)
               information)
           (PP on (NP Cuba)))
```

```
   (VP was
       (ADJP-PRD too outdated
                 (SBAR (WHNP-1 0)
                       (S (NP-SBJ *T*-1)
                          (VP to
                              (VP be
                                  (ADJP-PRD useful))))))))


(S (NP-SBJ Miss Schwarzkopf)
   (VP is
       (NP-PRD (NP (ADJP too great)
                   an artist)
               (SBAR (WHNP-1 0)
                     (S (NP-SBJ *T*-1)
                        (VP to
                            (VP need
                                (NP them))))))))
```

- No apparent gap.

```
(S (S (NP-SBJ-2 the risks)
      (VP were
          (ADJP-PRD-3 too high
                      (S *RNR*-1))))
   and
   (S (NP-SBJ=2 the potential payoff)
      (ADVP-PRD=3 too far
                  (PP in
                      (NP the future))
                  (S *RNR*-1)))
   (S-1 (NP-SBJ *)
        (VP to
            (VP justify
                (NP a higher offer)))))


(S (NP-SBJ It)
   (VP was
       (ADJP-PRD too late
                 (S (NP-SBJ *)
                    (VP to
                        (VP worry
                            (PP-CLR about
                                    (NP that))))))))
```

14.2.4 Purpose/reason clauses

Sentential purpose clauses. Normal infinitive purpose clauses (i.e., those without a gap) express a purpose/reason for the action described by the verb. They are simply labeled S-PRP and attached at VP level. The null subject is coindexed as appropriate.

```
(S (NP-SBJ-3 Mary)
   (VP took
       (NP a class)
       (S-PRP (NP-SBJ *-3)
              (VP to
                  (VP learn
                      (NP statistics))))))
```

```
(S (NP-SBJ-1 The figures)
   (VP were
       (VP adjusted
           (NP *-1)
           (S-PRP (NP-SBJ *)
                  (VP to
                      (VP remove
                          (NP (NP the effects)
                              (PP of
                                  (NP usual seasonal patterns)))))))))
```

Object and subject purpose clauses (OPCs and SPCs). Object and subject purpose clauses express a purpose for an object, as opposed to a purpose/reason for the action described by the sentence as a whole. They contain a gap and are usually bracketed as relative clauses, possibly with a -PRP tag on the SBAR or S. As noted, if the gap is not obvious, the clause may be bracketed as a sentential purpose/reason clause.

- Object.

```
(S (NP-SBJ We)
   (VP bought
       (NP (NP a broom)
           (SBAR-PRP (WHNP-1 0)
                     (S (NP-SBJ *)
                        (VP to
                            (VP sweep
                                (NP the floor)
                                (PP-MNR with
                                        (NP *T*-1)))))))))
```

- Subject.

```
(S (NP-SBJ John)
   (VP designed
       (NP (NP telescopes)
           (SBAR-PRP (WHNP-1 0)
                     (S (NP-SBJ *T*-1)
                        (VP to
                            (VP sit
                                (PP-LOC on
                                        (NP Kitt Peak)))))))))
```

SPCs are difficult to distinguish from infinitival semicomplements (see section 15 [Small Clauses], under "design-type verbs"), so the above sentence may receive the analysis below instead.

```
(S (NP-SBJ John)
   (VP designed
       (NP-3 telescopes)
       (S-CLR (NP-SBJ *-3)
              (VP to
                  (VP sit
                      (PP-LOC on (NP Kitt Peak)))))))
```

14.2.5 Complements of verbs

Indirect questions. Infinitival clauses headed by a *wh*-word are bracketed just as tensed complement clauses are, with a null subject.

```
(S (NP-SBJ-2 I)
   (VP do n't
       (VP know
           (SBAR (WHNP-1 what)
                 (S (NP-SBJ *-2)
                    (VP to
                        (VP do
                            (NP *T*-1)
                            (PP-CLR with
                                    (NP this sentence)))))))))
```

```
(S (NP-SBJ-2 I)
   (VP asked
       (NP the boss)
       (SBAR (WHNP-1 what)
             (S (NP-SBJ *-2)
                (VP to
                    (VP do
                        (NP *T*-1)))))))
```

Ditransitives. The Treebank recognizes only a very small number of verbs that allow an infinitive clause complement with a separate NP indirect object. The following list of ditransitives is based on [Quirk et al. 1985], section 16.63:

> *advise ask beg beseech challenge command counsel detail direct enjoin exhort forbid implore incite instruct invite order persuade pray promise remind request recommend teach tell urge*

By default, verbs not on this list are bracketed as monotransitives taking a single S complement. Note that verbs like *allow, authorize, bribe, encourage, force, inspire,* and *require* are not included here, and should be bracketed with a single S complement. (For ease of reference, much of this information is repeated in section 15 [Small Clauses]. See [Quirk et al. 1985], section 16.66, for more on the problem of distinguishing ditransitives from complex transitives.)

These infinitive clause complements have a null subject that is coindexed with the subject or indirect object of the verb, according to interpretation.

```
(S (NP-SBJ Ford)
   (VP persuaded
       (NP-1 Zaphod)
       (S (NP-SBJ *-1)
          (VP to
              (VP run
                  (PP-CLR for
                          (NP president)))))))
```

```
(S (NP-SBJ-1 Zaphod)
   (VP promised
       (NP Ford)
       (S (NP-SBJ *-1)
          (VP to
              (VP run
                  (PP-CLR for
                          (NP president)))))))
```

```
(S (NP-SBJ-1 Zaphod)
   (VP was
       (VP persuaded
           (NP-2 *-1)
           (S (NP-SBJ *-2)
              (VP to
                  (VP run
                      (PP-CLR for
                              (NP Galactic president))))))))
```

Monotransitives and complex transitives. All verbs in these classes are bracketed with a single clausal S complement. If there is no expressed subject, a null subject is added and coindexed with the logical antecedent. If the clause is introduced by *for*, *for* appears as a complementizer inside SBAR.

- With overt subject.

```
(S (NP-SBJ I)
   (VP believe
       (S (NP-SBJ it)
          (VP to
              (VP be
                  (NP-PRD a fair copy))))))
```

```
(S (NP-SBJ The following prompts)
   (VP allow
       (S (NP-SBJ you)
          (VP to
              (VP specify
                  (SBAR (WHADVP-1 how)
                        (S (NP-SBJ you)
                           (VP want
                               (S (NP-SBJ the printed output)
```

```
                                             (VP to
                                                 (VP look
                                                     (ADVP-MNR *T*-1)))))))))))
```

- With null subject.

```
(S (NP-SBJ-1 We)
   (VP decided
       (S (NP-SBJ *-1)
          (VP to
              (VP move)))))
```

This includes "raising" constructions:

```
(S (NP-SBJ-1 The Soviets)
   (VP are
       (ADVP widely)
       (VP believed
           (S (NP-SBJ *-1)
              (VP to
                  (VP need
                      (NP additional supplies)))))))
```

```
(S (NP-SBJ-1 the demon)
   (VP seems
       (S (NP-SBJ *-1)
          (VP to
              (VP have
                  (VP gone))))))
```

- after *be*

When *be* acts as a semi-modal, the subject of the infinitival clause is coindexed and the S should not receive a -PRD tag.

```
(S (NP-SBJ-1 You)
   (VP are
       (S (NP-SBJ *-1)
          (VP to
              (VP resign
                  (ADVP-TMP immediately))))))
```

However, if *be* is acting as a simple linking verb, the S complement is tagged -PRD and its subject is not coindexed.

```
(S (NP-SBJ His idea)
   (VP was
       (S-PRD (NP-SBJ *)
              (VP to
                  (VP leave
```

```
                     (ADVP-TMP (ADVP as soon)
                               (PP as
                                   (ADJP possible)))))))
```

- With *for*.

```
(S (NP-SBJ-1 I)
   (VP sat
       (S (NP-SBJ *-1)
          (VP waiting
              (SBAR for
                    (S (NP-SBJ Life)
                       (VP to
                           (VP (VP come
                                   (ADVP-CLR along))
                               and
                               (VP sweep
                                   (NP me)
                                   (PRT up))))))))))
```

- With extraposition.

```
(S (NP-SBJ I)
   (VP consider
       (S (NP-SBJ (NP it)
                  (S *EXP*-3))
          (VP to
              (VP be
                  (NP-PRD my job)
                  (S-3 (NP-SBJ *)
                       (VP to
                           (VP keep
                               (S (NP-SBJ-2 the customer)
                                  (VP (ADVP well)
                                      informed
                                      (NP *-2)))))))))))
```

Semi-complement clauses. Certain verbs (e.g., *use, design, hire, build*) are frequently followed by a closely related adjunct clause. However, they are not generally considered complements, since deleting the clause does not appreciably change the meaning of the verb. These clauses are frequently bracketed with -CLR, however, to indicate the middle ground that they occupy between argument and adjunct.

(See section 15 [Small Clauses] for examples and more information (under "*Use* and verbs like *design*") and also section 14.2.4 above (page 245, under "SPCs").)

Verbs like *cost* and *take*. These verbs appear in two different constructions that alternate with each other, as listed below. Possible approaches to bracketing are listed under each.

1. Infinitival semi-complement. Verbs like *cost* and *take* often introduce a (noun phrase) complement of duration or extent, followed by an infinitival semicomplement that contains a gap that is often interpreted as coreferential with the matrix subject:

The process / will take / as many as six months / to complete ___
The place / costs / nearly $2 million a year / to maintain ___

- Infinitival labeled S-CLR, gap not marked.

```
(S (NP-SBJ The process)
   (VP will
       (VP take
           (NP (QP as many as six)
               months)
           (S-CLR (NP-SBJ *)
                  (VP to
                      (VP complete))))))
```

- Infinitival labeled S-PRP, gap not marked.

```
(S (NP-SBJ The place)
   (VP costs
       (NP (NP (QP nearly $ 2 million) *U*)
           (NP-ADV a year))
       (S-PRP (NP-SBJ *)
              (VP to
                  (VP maintain)))))
```

- Infinitival labeled S-CLR, gap marked with *?* (ellipsed material).

```
(S (NP-SBJ The plant)
   (VP will
       (VP cost
           (NP (QP about 50 million) Canadian dollars)
           (S-CLR (NP-SBJ *)
                  (VP to
                      (VP build
                          (NP *?*)))))))
```

2. *it*-extraposition. These verbs also appear in a construction that looks like *it*-extraposition, though it is not necessarily bracketed that way (see section 17 [*It*-Extraposition] for more on the annotation of extraposition structures):

   It / will take / as many as six months / to complete the process.
   It / costs / nearly $2 million a year / to maintain the place.
   It / will take / Ford / 242 days / to sell off the current inventory.

   - Infinitival labeled S-PRP, infinitival null subject coindexed with matrix object.

```
(S (NP-SBJ it)
   (VP may
       (VP take
           (NP-2 us)
           (NP (QP six to nine) months)
           (S-PRP (NP-SBJ *-2)
                  (VP to
                      (VP find
                          (NP the money)))))))
```

- Extraposition structure, no coindexation.

```
(S (NP-SBJ (NP it)
           (S *EXP*-1))
   (VP may
       (VP take
           (NP us)
           (NP (QP six to nine) months)
           (S-1 (NP-SBJ *)
                (VP to
                    (VP find
                        (NP the money)))))))
```

15 Small Clauses and their near relatives

This section is concerned with the closely related complements of certain verbs. The verbs discussed here generally have a noun phrase complement that is the logical subject of a second complement that appears to be predicated of the first complement – a construction often referred to as a "small clause". However, since we cannot hope to capture all the subtleties of the small clause-type structures described in syntactic literature, we therefore annotate alike all complement pairs that show a predicative relationship. As a result, this section covers many constructions that are not "small clauses" in the technical sense, including constructions which upon closer inspection may in fact not merit the predicative analysis we give them here.

Note also that the policy described here turns out to be a bit too complex to follow entirely consistently. Users of the corpus should therefore expect a little roughness around the edges.

15.1 Bracketing

In general, non-finite clausal complements are labeled S. The "subject" of the clause is marked -SBJ, and the "predicate" is marked -PRD (unless the predicate is a VP, which never bears the -PRD tag).

```
(S (NP-SBJ I)
   (VP consider
       (S (NP-SBJ Kris)
          (NP-PRD a fool))))
```

If the verb is passive, the null passive object is shown as the subject of a clause:

```
(S (NP-SBJ-1 Kris)
   (VP is
       (VP considered
           (S (NP-SBJ *-1)
              (NP-PRD a fool))
           (PP by
               (NP-LGS most people)))))
```

Small clauses may be structurally distinguished from ordinary main clauses by the fact that they are immediately dominated by a VP and lack a tensed verb or modal (POS-tagged VBP, VBZ, VBD, or MD) in an S-level VP.

15.2 Small clause criteria

The main difficulty lies in distinguishing whether a particular verb has a clausal complement or two independent complements that are not so closely related (the latter exemplified by *give Calvin a comic book* or *persuade Hobbes to eat Calvin*, neither of which are given the small clause analysis). In all cases, for a pair of complements to be eligible for "small clause" bracketing, the NP complement must be the logical subject of the second complement.

In addition, it should meet certain requirements (in the form of syntactic and semantic tests which are specific to the nature of the predicate) as described in the following sections. The tests in a given section are only intended to apply to the sort of predicate that the section is concerned with. For example, the criteria for adjectival predicates do not apply to adverbial predicates, and vice-versa. The intention of this policy is to make bracketing decisions much faster and easier, by allowing the annotator to decide most cases by quick, easy rules, only resorting to Tests in a few cases. (Note that despite this intention, some annotators occasionally confused the sections and applied incorrect criteria to different predicates, especially among the non-verbal predicates. Again, users of the corpus should expect a little roughness around the edges.)

15.3 Overt subject clausal complements

15.3.1 Verbal predicates

1. *to*-infinitives.

   These are usually bracketed as a single S, with a small list of ditransitive exceptions (see below) and some variation in the case of verbs of the *use/design* type (see page 254). (See section 14 [Infinitives] for a slightly longer story on these.)

   (a) Monotransitives: S complement.
   Most verbs take a single S complement.

```
(S (NP-SBJ This)
   (VP does not
       (VP allow
           (S (NP-SBJ the mystery)
              (VP to
                  (VP invade
                      (NP us)))))))
```

```
(S (NP-SBJ he)
   (VP had
       (ADVP-TMP finally)
       (VP gotten
           (S (NP-SBJ (NP Chairman Bill Hollowell)
                      (PP of
                          (NP the committee)))
              (VP to
                  (VP set
                      (NP it)
                      (PP-CLR for
                              (NP (NP public hearing)
                                  (PP-TMP on
                                          (NP Feb. 22))))))))))
```

   (b) Ditransitives: NP + S complement.
   The following verbs are considered ditransitives and take an NP and an S complement. (Note that here we are only concerned with ditransitives that have **clausal** complements; see section 1 [Overview of Basic Clause Structure] and section 2 [Notation] for information about verbs (such as *give* or *compare*) that take two complements or semicomplements that are not clausal.)

   advise ask
   beg beseech
   challenge command counsel
   detail direct
   enjoin exhort
   forbid
   implore incite inform instruct invite
   order
   persuade pray promise
   remind request recommend
   teach tell
   urge

By DEFAULT, verbs not on this list are bracketed as monotransitives taking a single S complement. Note that verbs like *allow, authorize, bribe, encourage, force, inspire,* and *require* are not included here, and are bracketed with a single S complement.

```
(S (NP-SBJ He)
   (VP told
       (NP-1 me)
       (S (NP-SBJ *-1)
          (VP to
              (VP wake
                  (NP you))))))
```

(c) *Use* and verbs like *design*.

Verbs of this category are bracketed according to a system that grew out of annotator conventions and semi-official decisions that were followed to varying degrees, as described here.

- *Use*.
  Most annotators label the infinitival following *use* S-CLR:

```
(S (NP-SBJ-1 Miss McCullers)
   (VP uses
       (NP her
           (NX (NX (NP poet 's) instinct)
               and
               (NX (NP storyteller 's) skill)))
       (S-CLR (NP-SBJ *-1)
              (VP to
                  (VP reaffirm
                      (NP (NP her place)
                          (PP-LOC at
                                  (NP (NP the very top)
                                      (PP of
                                          (NP modern
                                              American writing))))))))))
```

  (But note that one annotator diverged significantly from this convention, being more likely to label infinitivals following active *use* S-PRP, with no coindexing of the null subject; and more likely to label infinitivals following passive *use* as S, with the null subject acting as the passive trace.)

- *Design*-type verbs (including *hire, build, write,* and others).
  This class consists of verbs that take a direct object and an optional infinitive whose subject generally receives its reference from the direct object. This type is not as common as *use*. The corpus reflects three basic approaches, none especially more common than the others.

  i. -PRP and/or -CLR.
     The infinitive may be tagged with -PRP or -CLR, with the null subject often coindexed with the direct object.

```
(S (NP-SBJ I)
   (VP designed
       (NP-7 the bomb)
       (S-CLR (NP-SBJ *-7)
              (VP to
                  (VP explode
                      (PP on
                          (NP contact)))))))
```

```
(S (NP-SBJ the vehicles)
   (VP were
       (VP built
           (NP-3 *-2)
           (S-PRP-CLR (NP-SBJ *-3)
                      (VP to
                          (VP meet
                              (NP U.S. government
                                  auto safety standards)))))))

(S (NP-SBJ the Senate Ethics Committee)
   (VP should
       (VP hire
           (NP-2 a special outside counsel)
           (S-CLR (NP-SBJ *-2)
                  (VP to
                      (VP conduct
                          (NP an investigation)))))))
```

ii. Simple monotransitive.
The infinitive may be bracketed as a simple monotransitive, especially in passive constructions.

```
(S (NP-SBJ he)
   (VP has
       (VP hired
           (S (NP-SBJ civil engineers)
              (VP to
                  (VP check
                      (PRT out)
                      (NP his own
                          (ADJP mildly damaged)
                          building)))))))

(S (NP-SBJ-1 the offering)
   (VP was
       (VP designed
           (S (NP-SBJ *-1)
              (VP to
                  (VP create
                      (NP (NP long-term incentives)
                          (PP for
                              (NP employees)))))))))
```

iii. Subject purpose clause.
The infinitive may be bracketed as a subject purpose clause (which might have -PRP on the S, -PRP on SBAR, or no -PRP at all).

```
(S (NP-SBJ John)
   (VP designed
       (NP (NP telescopes)
           (SBAR-PRP (WHNP-1 0)
                     (S (NP-SBJ *T*-1)
                        (VP to
```

```
                            (VP sit
                                (PP-LOC on
                                        (NP Kitt Peak))))))))

(S (NP-SBJ the company)
   (VP has
       (VP hired
           (NP (NP an investment banker)
               (SBAR (WHNP-2 0)
                     (S (NP-SBJ *T*-2)
                        (VP to
                            (VP help
                                (S (NP-SBJ it)
                                   (VP raise
                                       (NP additional cash)))))))))))
```

(d) Triple complements.

These are very rare. One common way of handling such constructions might be:

```
(S (PP-TMP In (NP 1607 and 1608))
   ,
   (NP-SBJ the English Muscovy Company)
   (VP had
       (VP sent
           (NP-1 him)
           (ADVP-DIR northward)
           (S-CLR (NP-SBJ *-1) <+ might have -PRP too +>
                  (VP to
                      (VP look
                          (PP-CLR for
                                  (NP (NP a route)
                                      (PP-LOC
                                       (PP over
                                           (NP the North Pole))
                                       or
                                       (PP across
                                           (NP (NP the top)
                                               (PP of
                                                   (NP Russia))))))))))))
```

2. Bare infinitives.

These are always bracketed as a single S. (see section 14 [Infinitives] list for details.)

```
(S (NP-SBJ The following)
   (VP helps
       (S (NP-SBJ you)
          (VP plan
              (PP-CLR for
                      (NP (NP future use)
                          (PP of
                              (NP the system monitor))))))))
```

3. Present participles.

These are bracketed as a single S. (See also section 13 [Gerunds and Participles].)

```
(S (NP-SBJ You)
   (VP 'd
       (VP see
           (S (NP-SBJ her)
              (VP correcting
                  (NP homework)
                  (PP-LOC in
                          (NP the stands))
                  (PP-LOC at
                          (NP a football game)))))))
```

```
(S (ADVP Lastly)
   ,
   (NP-SBJ-1 governmental and private planners)
   (VP will
       (PP-TMP at (NP this stage))
       (VP begin
           (S (NP-SBJ *-1)
              (VP to
                  (VP see
                      (S (NP-SBJ large capital requirements)
                         (VP looming
                             (ADVP-LOC-CLR ahead)))))))))
```

(Note that this could also be interpreted as *"requirements* **which are** *looming"*, in which case it should be bracketed as a reduced relative. See section 13 [Gerunds and Participles] for details on the bracketing of reduced relatives.)

4. Past participles.

   These are bracketed as a single S. The gap contains a passive trace coindexed with the clausal subject.

```
(S (NP-SBJ I)
   (VP want
       (S (NP-SBJ-1 (NP the room)
                    (PP-LOC in
                            (NP the attic)))
          (VP prepared
              (NP *-1)
              (PP-BNF for
                      (NP him))))))
```

```
(S (NP-SBJ (NP the man)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP submits
                        (NP a manuscript)
                        (PP-CLR to (NP a publisher))))))
   (VP will
       (VP find
```

```
              (S (NP-SBJ-3 himself)
                 (VP reviewed
                     (NP *-3)
                     (SBAR-TMP before
                               (S (NP-SBJ-2 he)
                                  (VP is (VP accepted (NP *-2)))))))))
```

```
(S (NP-SBJ I)
   (VP had
       (VP had
           (S (NP-SBJ-1 my name)
              (VP taken
                  (NP *-1)
                  (PP-DIR out
                          (PP of
                              (NP the telephone book)))))))
```

15.3.2 Adjectival predicates

These are generally given a clausal analysis. Verbs that receive the clausal analysis for their adjectival predicates include but are not limited to *hold, keep, leave; call, report; like, prefer, want; believe, consider, find, imagine, presume, think; drive, get, make; certify, declare, proclaim.* (This list is based on [Quirk et al. 1985], section 16.44.)

```
(S (NP-SBJ People)
   (VP find
       (S (NP-SBJ this statement)
          (ADJP-PRD utterly impenetrable))))
```

1. With *as*.

   Adjectival predicates introduced by *as* are labeled PP-CLR, with no small clause.

   ```
   (S (NP-SBJ They)
      (VP do not
          (VP consider
              (NP (NP themselves)
                  and
                  (NP their plight))
              (PP-CLR as
                      (ADJP statistical)))))
   ```

2. Resultatives.

   A few adjectival "predicates" are more adverbial than predicative and are not given the clausal analysis. These are extremely rare (12 in the entire WSJ corpus). This bracketing is used only when eliminating the putative predicate causes no substantial change in the meaning of the rest of the sentence.

   That is, if you:

   lick the platter clean
   eat fish uncooked

paint the town red
shoot the man dead
pound the nail flat

you also:

lick the platter
eat fish
paint the town
shoot the man
pound the nail

However, if you:

drive your parents crazy
get your brother in trouble
make your sister happy
declare the manual unwritable

you generally don't:

drive your parents
get your brother
make your sister
declare the manual

Thus the second set of verbs take a simple small clause analysis. The first set, on the other hand, get the more complicated resultative structure:

```
(S (NP-SBJ They)
   (VP painted
       (NP-1 the apartment)
       (S-CLR (NP-SBJ *-1)
              (ADJP-PRD orange
                        ,
                        pink
                        and
                        white))
       ,
       (PP according
           (PP to
               (NP her instructions)))))


(S (NP-SBJ (NP the government 's)
           action)
   (VP caught
       (NP-2 Jaguar management)
       (S-ADV (NP-SBJ *-2)
              (ADJP-PRD flat-footed))))


(S (NP-SBJ She)
   (VP had
       (VP (VP raised
               (NP a calf))
```

```
                ,
                (VP grown
                    (NP-1 it)
                    (S-CLR (NP-SBJ *-1)
                           (ADJP-PRD beef-fat))))))
```

3. Pseudo-adjectives. Phrases with adjectival meaning and predicate adjective distribution that are labeled something other than ADJP also receive a clausal analysis.

```
(S (NP-SBJ John)
   (VP got
       (S (NP-SBJ Bob)
          (PP-PRD in
                  (NP trouble)))))


(S (NP-SBJ Sam)
   (VP wanted
       (S (NP-SBJ the whole house)
          (PP-PRD in
                  (NP order))
          (SBAR-TMP before
                    (S (NP-SBJ he)
                       (VP came
                           (ADVP-DIR downstairs)))))))
```

15.3.3 Nominal predicates

Verbs with two NP complements receive either a small clause analysis:

```
(S (NP-SBJ The late
           (NAC Secretary (PP of (NP State)))
           John Foster Dulles)
   (VP considered
       (S (NP-SBJ the 1954 Geneva agreement)
          (NP-PRD (NP a specimen)
                  (PP of (NP appeasement))))))
```

or a double object analysis:

```
(S (NP-SBJ His bel canto style)
   (VP gave
       (NP the performance)
       (NP a special distinction)))
```

Small clauses are distinguished by an equative relationship between the first complement (NP1) and the second (NP2). That is, two NP complements are given the clausal analysis when it makes sense to say "NP1 is NP2". In almost all cases where a small clause bracketing was conceivable, the annotator made the NPs into a small clause. Therefore, virtually all NP pairs that are directly under a VP and not within a complement clause (and also not tagged with -CLR, -TMP, etc.) are indirect object/direct object pairs.

Verbs that can take this kind of Small Clause include *hold, keep, leave, call, pronounce; wish; believe, consider, find, imagine, think; appoint, elect, make, vote; certify, christen, declare, name,* among others.

For example:

- Small Clause

```
(S (NP-SBJ House Speaker Sam Rayburn)
   (VP called
       (S (NP-SBJ the Kennedy program)
          ``
          (NP-PRD a (ADJP mighty fine) thing)
          ''))))
```

```
(S (NP-SBJ That)
   (VP makes
       (S (NP-SBJ them)
          (NP-PRD a reasonable option))))
```

```
(S (NP-SBJ-29 (NP John A. Conlon Jr.)
              ,
              (NP 45)
              ,)
   (VP was
       (VP named
           (S (NP-SBJ *-29)
              (NP-PRD (NP a managing director)
                      (PP-LOC at
                              (NP this investment-banking company)))))))
```

- Double object

```
(S (NP-SBJ I)
   (VP permitted
       (NP the invading car)
       (NP free access)))
```

15.3.4 Adverbial predicates

Most ADVP and PP complements that express direction or location are **not** shown as small clauses, and are instead tagged with -CLR or -LOC-CLR (with the exception of *put,* whose locative complement gets -PUT). These verbs include *put, get, send, drive, show, keep, leave,* among others.

```
(S (NP-SBJ We)
   (VP put
       (NP the proposal)
       (PP-PUT on (NP the table))))
```

```
(S (NP-SBJ We)
   (VP left
       (NP the cat)
       (PP-LOC-CLR at (NP home))))
```

There are, however, a few (about 100 tokens) adverbial small clauses in the corpus. These include cases where the "adverbial" is actually adjectival, but receives a bracket label such as PP (see "pseudo-adjectives" in section 3 (page 260) for more information). They also include cases where the annotator had an strong intuition of small-clauseness, misunderstood the policy, or simply made an error. This is most likely to happen with verbs like *get, keep,* and *leave,* which frequently appear with adjectival predicates.

```
(S (NP-SBJ The doctor)
   (VP wanted
       (S (NP-SBJ him)
          (PP-LOC-PRD in (NP a hospital)))))


(S (NP-SBJ they)
   ``
   (VP could
       (VP get
           (S (NP-SBJ some roadblocks)
              (PP-PRD out
                      (PP of
                          (NP the way)))))))
```

With *as*: Predicates introduced with *as* are labeled with PP-CLR, without a small clause:

```
(VP consider
    (NP Spain , Portugal and Greece)
    (PP-CLR as
            (NP possible manufacturing sites)))


(VP listed
    (NP (NP the mayor 's)
        occupation)
    (PP-CLR as
            `` (NP attorney) ''))
```

15.3.5 Particle predicates

Particles are not treated as clausal predicates, although some particle verbs appear with constructions that resemble small clauses. For example, in *turn the light off,* "the light" is "off" as a result of the action described by the verb. We do not, however, bracket these as small clauses.

```
(S (NP-SBJ *)
   (VP turn
       (NP the light)
       (PRT off)))
```

15.4 Null subject clausal complements

15.4.1 Active

1. *to*-infinitive complement.

If the verb is active, and there is a single infinitival complement (one which is not a purpose clause), the complement S is bracketed with a null subject. In general, the null subject should also be coindexed with the subject of the main clause. However, if the sentential subject is not interpreted as the subject of the lower clause, there is no coindexation.

```
(S (NP-SBJ-1 I)
   (VP would
       (VP like
           (S (NP-SBJ *-1)
              (VP to
                  (VP think
                      (PP-CLR of
                              (NP (NP a way)
                                  (SBAR
                                   (WHADVP-2 0)
                                   (S (NP-SBJ *)
                                      (VP to
                                          (VP make
                                              (NP a little extra money)
                                              (ADVP-MNR *T*-2))))))))))))
```

```
(S But
   ,
   (INTJ alas)
   ,
   (NP-SBJ-1 the authenticity)
   (VP seems
       (S (NP-SBJ *-1)
          (VP to
              (VP stop
                  (PP-LOC-CLR at
                              (NP (NP the set 's) edge)))))))
```

2. *-ing* participial clauses (includes "serial verbs").

   These are also labeled S and given null subjects, coindexed as appropriate. (See section 13 [Gerunds and Participles] for more information on the bracketing of participial complements of VP.)

```
(S (NP-SBJ-1 I)
   (VP 'd
       (ADVP rather)
       (VP (VP keep
               (S (NP-SBJ *-1)
                  (VP bailing)))
           --
           or
           (VP sink))))
```

```
(S (ADVP Characteristically)
   ,
   (NP-SBJ-1 Trevelyan)
```

```
(VP enjoyed
    (S (NP-SBJ *-1)
       (VP writing
           (NP the work)))))
```

15.4.2 Passive

1. Small clause complements (as defined in section 15.3 above).

   Passives of verbs that take single clausal complements are annotated with the (NP *) passive trace in subject position of the small clause. Examples are given here for verbal, adjectival, and nominal predicates.

   - Verbal.

```
(S (NP-SBJ-1 Professional responsibility)
   (VP is
       (VP seen
           (S (NP-SBJ *-1)
              (VP to
                  (VP consist
                      (ADVP largely)
                      (PP-CLR in
                              (S-NOM (NP-SBJ *)
                                     (VP serving
                                         (NP (NP the wishes)
                                             (PP of (NP the client)))
                                         (UCP-MNR (ADVP fairly)
                                                  and
                                                  (PP in
                                                      (NP an efficient
                                                          manner))))))))))))
```

```
(S (NP-SBJ-1 (NP the tall figure)
             (PP with
                 (NP the rifle and field glasses)))
   (VP had
       (VP been
           (VP seen
               (S (NP-SBJ *-1)
                  (VP riding
                      (NP-DIR that way)))))))
```

   - Adjectival.

```
(S-3 (NP-SBJ-1 Such vital information)
     (PRN ,
          (S (NP-SBJ he)
             (VP said
                 (SBAR 0
                       (S *T*-3))))
          ,)
     (VP has
         (S (NP-SBJ-2 *-1)
```

```
          (VP to
              (VP be
                  (VP made
                      (S (NP-SBJ *-2)
                         (ADJP-PRD available
                                   (PP to
                                       (NP the public))))
                      (UCP-TMP (ADVP frequently)
                               and
                               (PP at (NP regular intervals)))
                      (SBAR-PRP for
                                (S (NP-SBJ residents)
                                   (VP to
                                       (VP know))))))))))
```

- Nominal.

```
(S And
   (NP-SBJ-1 he)
   (VP was
       n't
       (ADVP really)
       (VP elected
           (S (NP-SBJ *-1)
              (NP-PRD (NP treasurer)
                      (PP of (NP the science club)))))))
```

```
(NP (NP a factor)
    (SBAR (WHNP-1 which)
          (S (NP-SBJ *T*-1)
             (VP has
                 (VP caused
                     (S (NP-SBJ the Citizens Group)
                        (VP to
                            (VP obtain
                                (NP signatures)
                                (PP under
                                    (SBAR-NOM
                                     (WHNP-2 what)
                                     (S (NP-3 *T*-2)
                                        (VP were
                                            (VP termed
                                                (S (NP-SBJ *-3)
                                                   ``
                                                   (NP-PRD false pretenses)
                                                   ''))))))))))))))
```

2. Double complement passives.

   Verbs that take a double complement when not passive receive the same analysis when in the passive, with the same sort of coindexing.

   - Ditransitive infinitive (cf. section 15.3.1).

Verbs such as *persuade* in the passive are annotated with two null elements: a passive trace coindexed with the structural subject and a null subject in the infinitive coindexed with the passive object.

```
(S (NP-SBJ-1 Zaphod)
   (VP was
       (VP persuaded
           (NP-2 *-1)
           (S (NP-SBJ *-2)
              (VP to
                  (VP run
                      (PP-CLR for (NP Galactic president))))))))
```

- Adverbial adjective/Resultative (cf. section 15.3.2)

```
(S (NP-SBJ-1 Fish)
   (VP are
       (ADVP-TMP sometimes)
       (VP eaten
           (NP-2 *-1)
           (S-CLR (NP-SBJ *-2)
                  (ADJP-PRD uncooked)))))
```

15.5 Special problems

15.5.1 Heavy shift

When the NP is "heavy" and therefore moved past the predicate, the clause is bracketed S and the different word order is represented by the -SBJ tag on the subject.

```
(S (NP-SBJ I)
   (VP (VP lifted
           (NP the lines)
           (PP-CLR from (NP the dust)))
       and
       (VP found
           (S (VP hitched
                  (NP *-2)
                  (PP-CLR to (NP that plow)))
              (NP-SBJ-2 (NP the finest team)
                        (SBAR (WHNP-1 0)
                              (S (NP-SBJ I)
                                 (ADVP-TMP ever)
                                 (VP held
                                     (NP a rein)
                                     (PP-CLR on
                                             (NP *T*-1))))))))))
```

```
(S (NP-SBJ The Supreme Court)
   (VP let
       (S (VP stand)
          (NP-SBJ (NP a New York court 's)
                  ruling
```

```
                    (SBAR that ...)))))
```

(Note in some cases annotators labeled these clauses SINV, even though this is slightly inconsistent with the definition of SINV as "S containing tensed verb or modal preceding clause subject". Since small clauses don't contain a tensed verb or modal, by definition they should not be labeled SINV.)

15.5.2 Coordination

Small clauses, like everything else, are coordinated as low as possible.

```
(S (NP-SBJ The cow)
   (VP kept
       (S (S (NP-SBJ her eyes)
             (ADJP-PRD open))
          and
          (S (NP-SBJ her mind)
             (PP-PRD on (NP her business))))))


(S (NP-SBJ (NP any one)
           (PP of
               (NP these spiced meats)))
   (VP (VP makes
           (S-4 (NP-SBJ a man)
                (NP-PRD a cook)))
       ,
       and
       (VP (ADVP-TMP sometimes)
           (S=4 (NP-SBJ a meal)
                (NP-PRD a feast)))))
```

16 Clefts

16.1 *It*-clefts (or "true" clefts)

Cleft sentences contain the non-referential subject pronoun *it*, followed by a form of the verb *be*, followed by the clefted portion, followed by a final clause with gap (e.g., *It / is / the clumsy child / who sustains the worst injuries*).

16.1.1 Declarative *it*-clefts

Declarative *it*-clefts are labeled S-CLF; *it* receives the -SBJ tag as the surface subject; the clefted part (or focus) is tagged -PRD as the complement of *be*; the final clause (SBAR) is attached at VP level and is not marked with a dash tag.

The adverbial function tags (-LOC, -MNR, -PRP, -TMP, etc.) are used only on the trace in the subordinate clause, not in the clefted part of the sentence, which may only be tagged -PRD.

1. Subject as focus:

```
(S-CLF (NP-SBJ It)
       (VP is
           (NP-PRD the clumsy child)
           (SBAR (WHNP-1 who)
                 (S (NP-SBJ *T*-1)
                    (VP sustains
                        (NP the worst injuries))))))
```

2. Direct object as focus:

```
(S-CLF (NP-SBJ It)
       (VP is
           (NP-PRD this conclusion)
           (SBAR (WHNP-1 that)
                 (S (NP-SBJ we)
                    (VP challenge
                        (NP *T*-1))))))
```

3. Manner as focus:

```
(S-CLF (NP-SBJ It)
       (VP was
           (ADVP-PRD slowly)
           (SBAR (WHADVP-1 that)
                 (S (NP-SBJ Kris)
                    (VP opened
                        (NP the door)
                        (ADVP-MNR *T*-1))))))
```

4. Time as focus:

```
(S-CLF (NP-SBJ It)
       (VP is
           (ADVP-PRD then)
           (SBAR (WHADVP-1 that)
                 (S (NP-SBJ-2 young queens)
                    (VP begin
                        (S (NP-SBJ *-2)
                           (VP to
                               (VP appear)))
                        (ADVP-TMP *T*-1))))))
```

5. Location as focus:

```
(S-CLF (NP-SBJ It)
       (VP was
           (PP-PRD at
                   (NP the bus stop))
           (SBAR (WHADVP-1 that)
                 (S (NP-SBJ we)
                    (VP met
                        (ADVP-LOC *T*-1))))))
```

6. Purpose/reason as focus:

```
(S-CLF (NP-SBJ It)
       (VP was
           (SBAR-PRD because
                     (S (NP-SBJ he)
                        (VP got
                            (ADJP-PRD sick))))
           (SBAR (WHADVP-1 that)
                 (S (NP-SBJ we)
                    (VP cancelled
                        (NP the meeting)
                        (ADVP-PRP *T*-1))))))
```

16.1.2 Interrogative *it*-clefts

Interrogative cleft sentences are labeled SQ-CLF (see section 1 [Overview of Basic Clause Structure] for the treatment of SQs). The rest of the sentence is annotated as described above for declarative clefts.

```
(SQ-CLF Was
        (NP-SBJ it)
        (NP-PRD John)
        (SBAR (WHNP-1 who)
              (S (NP-SBJ *T*-1)
                 (VP came
                     (PP-DIR to
                             (NP the party))
                     (PP-MNR in
                             (NP a dress)))))
        ?)
```

16.1.3 Inversion in *it*-clefts

A cleft sentence that appears inside an SINV is labeled SINV-CLF and the postposed clause is attached at VP level, as with other types of clefts described in this section. In this case, the inverted cleft functions as a conditional clause and is thus labeled SBAR-ADV (see section 1 [Overview of Basic Clause Structure] for more on the treatment of inverted conditionals):

```
(S (SBAR-ADV (SINV-CLF Had
                       (NP-SBJ it)
                       (VP been
                           (NP-PRD (NP Wonder Woman)
                                   (CONJP rather than)
                                   (NP Aquaman))
                           (SBAR (WHNP-1 who)
                                 (S (NP-SBJ *T*-1)
                                    (VP was
                                        (VP flying
                                            (NP the plane))))))))
   (NP-SBJ it)
   (VP might
       not
       (VP have
           (VP crashed))))
```

16.1.4 Gapping across cleft sentences

Gapping across cleft sentences proceeds in the same way as with other instances of gapping.

```
(S-CLF (PP-TMP In
               (NP the past))
       ,
       (NP-SBJ it)
       (VP has
           (VP (VP been
                   (NP-PRD-2 the husband)
                   (SBAR (WHNP-1 who)
                         (S (NP-SBJ *T*-1)
                            (VP has
                                (VP been
                                    (ADJP-PRD-3 dominant))))))
               and
               (VP (NP-PRD=2 the wife)
                   (ADJP-PRD=3 passive)))))
```

]

16.2 *Wh*-clefts (or "pseudo" clefts)

Wh-clefts do not receive special treatment in the corpus. They contain a free/headless relative, followed by a form of the verb *be*, followed by a predicate (e.g., *What sustained the worst injuries / was / the car*). (See section 9 [*WH*-Phrases] on free relatives.)

17 *It*-Extraposition

17.1 *It*-extraposition from subject position

When a clausal subject is postposed, expletive *it* appears in the structural subject position. Characteristic of *it*-extraposition is that the final clause can replace *it*: *It is a pleasure to teach her* → *To teach her is a pleasure.*

17.1.1 Declaratives with *it*-extraposition

Declaratives with *it*-extraposition are analyzed as *it* + *be* + predicate + logical subject. Sentences that contain extraposed clausal subjects are labeled S. The final clause is attached at VP level and adjoined to the *it* with *EXP*-attach, and the NP containing the two is tagged -SBJ as the surface subject. Like the SBARs in *it*-clefts, and unlike all other non-complement SBARs in VP, the extraposed clauses do not have adverbial dash tags. As usual, the complement of *be* bears the -PRD tag. Note that the -LGS tag is not used in this case, but only for the logical subjects of passive sentences (see section 1 [Overview of Basic Clause Structure] and section 4 [Null Elements] on -LGS).

```
(S (NP-SBJ (NP It)
           (S *EXP*-1))
   (VP is
       (NP-PRD a pleasure)
       (S-1 (NP-SBJ *)
            (VP to
                (VP teach
                    (NP her))))))
```

```
(S (NP-SBJ (NP It)
           (SBAR *EXP*-1))
   (VP does
       n't
       (VP matter
           (SBAR-1 (WHNP-2 what)
                   (S (NP-SBJ you)
                      (VP do
                          (NP *T*-2)))))))
```

```
(S (NP-SBJ (NP it)
           (S *EXP*-3))
   (VP behooves
       (NP the hospital management)
       (S-3 (NP-SBJ *)
            (VP to
                (VP do
                    (NP some
                        (ADJP mighty careful)
                        planning))))))
```

```
(S (NP-SBJ (NP It)
           (S *EXP*-2))
   (VP 's
```

```
(ADJP-PRD (ADJP easier)
          (SBAR *ICH*-1))
(S-2 (NP-SBJ *)
     (VP to
         (VP get
             (ADJP-PRD worse))))
(SBAR-1 than
        (FRAG (ADJP-PRD better)))
(PP-LOC in
        (NP this game))))
```

17.1.2 Interrogatives with *it*-extraposition

Interrogatives with *it*-extraposition are labeled SQ (see section 1 [Overview of Basic Clause Structure] for the treatment of SQ). The rest of the sentence is annotated as described above for the declarative version.

```
(SQ Is
    (NP-SBJ (NP it)
            (SBAR *EXP*-1))
    (ADJP-PRD possible)
    (SBAR-1 that
            (S (NP-SBJ he)
               (VP killed
                   (NP her))))
    ?)
```

17.1.3 Inversion with *it*-extraposition

Extraposition may occur within SINV (see section 1 [Overview of Basic Clause Structure] for the treatment of SINV):

```
(SINV (PP-LOC Under
              (NP no circumstances))
      would
      (NP-SBJ (NP it)
              (S *EXP*-1))
      (ADVP-TMP ever)
      (VP be
          (ADJP-PRD acceptable)
          (S-1 (NP-SBJ *)
               (VP to
                   (VP permit
                       (NP (NP the termination)
                           (PP of
                               (NP the human race))))))))
```

17.1.4 Exclamatives with *it*-extraposition

Exlamatives that contain *it*-extraposition are labeled SBAR, and contain the *wh*-phrase, followed by the expletive subject, followed by *be*, followed by a coindexed trace in the position of the gap, followed by the extraposed clause (e.g., *How bizarre / it / is / that they eat bugs!*).

```
(SBAR (WHADJP-1 How bizarre)
      (S (NP-SBJ (NP it)
                 (SBAR *EXP*-2))
         (VP is
             (ADJP-PRD *T*-1)
             (SBAR-2 that
                     (S (NP-SBJ they)
                        (VP eat
                            (NP bugs))))))
      !)
```

17.1.5 Ambiguity with *it*-extraposition

A possible ambiguity arises when the final clause is headed by *for*.

Infinitival with overt subject. In some cases, the entire *for* clause is extraposed and therefore can be fronted to replace *it* (e.g., *It was impossible for anyone to escape* → *For anyone to escape was impossible*). Treating the entire *for* clause as extraposed represents the DEFAULT option.

```
(S (NP-SBJ (NP It)
           (SBAR *EXP*-1))
   (VP was
       (ADJP-PRD impossible)
       (SBAR-1 for
               (S (NP-SBJ anyone)
                  (VP to
                      (VP escape))))))
```

PP + infinitival. In other cases, it is only the infinitive (with null subject) that is extraposed and *for* heads an independent PP (e.g., *It is difficult for Willie to resist chocolate* → *To resist chocolate is difficult for Willie* ↛ **For Willie to resist chocolate is difficult*).

```
(S (NP-SBJ (NP It)
           (S *EXP*-2))
   (VP is
       (ADJP-PRD difficult)
       (PP for
           (NP Willie))
       (S-2 (NP-SBJ *)
            (VP to
                (VP resist
                    (NP chocolate))))))


(S (NP-SBJ (NP It)
           (S *EXP*-1))
   (VP was
       (ADJP-PRD easy)
       (PP for
           (NP the psalmist))
       (S-1 (NP-SBJ *)
            (VP to
                (VP sing
```

```
                        (NP them)
                        (PP-TMP in
                                (NP his day)))))))
```

17.1.6 Gapping with *it*-extraposition

Gapping across sentences with *it*-extraposition proceeds in the same way as with other instances of gapping.

```
(S (S (NP-SBJ It)
      (VP (VP is
              (ADJP-PRD-1 nice)
              (PP for
                  (NP teachers))
              (S-2 (NP-SBJ *)
                   (VP to
                       (VP think
                           (SBAR 0
                                 (S (NP-SBJ-3 they)
                                    (VP are
                                        (VP engaged
                                            (NP *-3)
                                            (PP-CLR in
                                                    ``
                                                    (NP personality development)
                                                    '')))))))))
          and
          (VP (ADVP-PRD=1 even nicer)
              (S=2 (NP-SBJ *)
                   (VP to
                       (VP minimize
                           (NP (NP those irksome tests)
                               (PP with
                                   (NP (ADJP often disappointing)
                                       results))))))))))
```

17.1.7 *It*-extraposition in small clauses

When *it*-extraposition occurs from the subject position of a small clause, it is treated in a manner similar to extraposition of the clausal subject, as described in section 17.1.1, with the exception that in this case the extraposed clause is put at S-level rather than VP-level (which in the small clause does not exist). For example: *I find it annoying that you constantly interrupt me when I am speaking* → *I find that you constantly interrupt me when I am speaking annoying*.

```
(S (NP-SBJ I)
   (VP find
       (S (NP-SBJ (NP it)
                  (SBAR *EXP*-1))
          (ADJP-PRD annoying)
          (SBAR-1 that
                  (S (NP-SBJ you)
                     (ADVP-TMP constantly)
                     (VP interrupt
                         (NP me)
```

```
                              (SBAR-TMP (WHADVP-2 when)
                                        (S (NP-SBJ I)
                                           (VP am
                                               (VP speaking
                                                   (ADVP-TMP *T*-2)))))))))))
```

```
(S (NP-SBJ We)
   (VP made
       (S (NP-SBJ (NP it)
                  (S *EXP*-1))
          (NP-PRD our objective)
          (S-1 (NP-SBJ *)
               (VP to
                   (VP finish
                       (NP the projects)
                       (PP-MNR in
                               (NP a timely fashion))))))))
```

17.2 *It*-extraposition from object position

There is no established policy for the annotation of these constructions, and *it*-extraposition from object position is quite rare in the corpus. This is how one example was done:

```
(S But
   (NP-SBJ I)
   (VP liked
       (NP it)
       (SBAR that
             (S (NP-SBJ they)
                (VP 'd
                    (VP brought
                        (NP their children)))))))
```

Most other cases of putative object *it*-extraposition are analyzed in the corpus as extraposition from the subject position of a small clause (see section 17.1.7).

18 Subject-Raising Predicates

Subject-raising predicates are those wherein the subject of an embedded clause can (optionally) be raised to the subject position of the matrix clause (e.g., *It appears that John is sick* → *John(i) appears PRO(i) to be sick*). This section describes the annotation of *unraised* raising predicates only. **Raised** subject-raising predicates are discussed in section 14 [Infinitives].

The *that* clause is bracketed as a complement of the subject-raising verb or adjective, and the *it* in subject position is labeled -SBJ.

18.1 Active verbs

The subject-raising verbs include *appear, be, chance, happen, seem,* etc.

```
(S (NP-SBJ It)
   (VP seems
       (SBAR that
             (S (NP-SBJ he)
                (VP was
                    (VP lying))))))
```

```
(S (NP-SBJ It)
   (VP may
       (VP be
           (SBAR-PRD that
                     (S (NP-SBJ I)
                        (VP was
                            (ADJP-PRD wrong)))))))
```

18.2 Passive verbs

The subject-raising passive verbs include *add, assume, believe, claim, decide, find, grant, hope, know, observe, prove, remember, say, see, show, suggest,* etc. Unlike with *it*-extraposition, here the *that* clause is an argument of the verb even though it can be fronted. There is thus no passive trace. Note that in their active forms all of these verbs take a *that* clause complement, even without the additional NP complement.

```
(S (NP-SBJ It)
   (VP was
       (ADVP-MNR widely)
       (VP believed
           (SBAR that
                 (S (NP-SBJ the world)
                    (VP was
                        (ADJP-PRD flat)))))))
```

```
(S (NP-SBJ It)
   (VP has
       (VP been
           (VP claimed
               (SBAR that
                     (S (NP-SBJ money)
                        (VP is
```

```
                              (NP-PRD (NP the root)
                                      (PP of
                                          (NP all evil))))))))))
```

A note on variation: these are easy to confuse with *it*-extraposition, and some examples in the corpus are bracketed as extraposition, with a passive trace.

```
(S (NP-SBJ-1 (NP It)
             (SBAR *EXP*-2))
   (VP was
       (ADVP-MNR widely)
       (VP believed
           (NP *-1)
           (SBAR-2 that
                   (S (NP-SBJ the world)
                      (VP was
                          (ADJP-PRD flat)))))))
```

18.3 Adjectives

The subject-raising adjectives include *certain, likely,* etc. (See section 14 [Infinitives] for more about this construction and its *raised* variant.) Note that these also are easily confused with *it*-extraposition.

```
(S (NP-SBJ It)
   (VP is
       (ADJP-PRD likely
                 (SBAR that
                       (S (NP-SBJ I)
                          (VP will
                              (VP be
                                  (PP-PRD on
                                          (NP time)))))))))

(S (NP-SBJ It)
   (VP is
       (ADJP-PRD certain
                 (SBAR that
                       (S (NP-SBJ he)
                          (VP will
                              (VP win)))))))
```

18.4 Subject-raising predicates in small clauses

Subject-raising predicates may also appear in small clauses.

```
(S (NP-SBJ *)
   (VP Let
       (S (NP-SBJ it)
          (VP be
              (VP granted
                  (ADVP then)
                  (SBAR that
                        (S (NP-SBJ-1 (NP the theological differences)
```

```
                                        (PP-LOC in
                                                (NP this area))
                                        (PP between
                                            (NP (NP Protestants)
                                                and
                                                (NP Roman Catholics))))
                              (VP appear
                                  (S (NP-SBJ *-1)
                                     (VP to
                                         (VP be
                                             (ADJP-PRD irreconcilable)))))))))))
```

18.5 Inversion with subject-raising predicates

The following are examples of subject-raising predicates in SINV:

```
(SINV (ADVP-LOC Nowhere)
      is
      (NP-SBJ it)
      (VP proven
          (SBAR that
                (S (NP-SBJ-1 you)
                   (VP have
                       (S (NP-SBJ *-1)
                          (VP to
                              (VP be
                                  (ADJP-PRD rich))))
                       (S-PRP (NP-SBJ *-1)
                              (VP to
                                  (VP be
                                      (ADJP-PRD happy)))))))))


(SINV (PP in
          (NP no way))
      is
      (NP-SBJ it)
      (ADJP-PRD likely
                (SBAR that
                      (S (NP-SBJ we)
                         (VP will
                             (VP solve
                                 (NP (NP all the problems)
                                     (PP of
                                         (NP syntax))))))))
```

19 Weather *it* and Referential *it*

Weather *it* and referential *it* receive no special annotation.

```
(S (NP-SBJ It)
   (VP 's
       (VP raining)))
```

Instances of referential *it* include uses such as the following:

```
(S (NP-SBJ It)
   (VP 'll
       (ADVP probably)
       (VP be
           (NP-TMP-PRD (QP at least an)
                       hour
                       (QP or two))
           (SBAR-TMP before
                     (S (NP-SBJ I)
                        (VP can
                            (VP check
                                (ADVP-CLR back)
                                (PP-CLR with
                                        (NP you)))))))))
```

```
(S (NP-SBJ It)
   (VP is
       (NP-PRD (NP time)
               (SBAR (WHADVP-2 0)
                     for
                     (S (NP-SBJ you)
                        (VP to
                            (VP go
                                (ADVP-TMP *T*-2))))))))
```

20 Existential *there*

Note that the ATIS sample included in this release uses an extraposition treatment of these cases, while the WSJ corpus does not.

20.1 ATIS bracketing

This construction appears only in questions in the ATIS sample.

```
(SQ Are
    (NP-SBJ (NP there)
            (NP *EXP*-1))
    (NP-1 (NP any flights)
          (VP arriving
              (PP-TMP after
                      (NP eleven a.m)))))
```

```
(SBARQ (WHNP-1 (WHADJP How many) stops)
       (SQ are
           (NP-SBJ (NP there)
                   (NP *EXP*-2))
           (NP-2 *T*-1)))
```

```
(SBARQ (WHNP-1 (WHNP What flights)
               (PP-DIR *ICH*-3)
               (PP-DIR *ICH*-4)
               (PP-TMP *ICH*-5))
       (SQ are
           (NP-SBJ (NP there)
                   (NP *EXP*-2))
           (NP-2 *T*-1)
           (PP-DIR-3 from
                     (NP Milwaukee))
           (PP-DIR-4 to
                     (NP Phoenix))
           (PP-TMP-5 on
                     (NP Saturday))))
```

20.2 WSJ bracketing

```
(S (NP-SBJ There)
   (VP is
       (NP-PRD a lovely statue)
       (PP-LOC in
               (NP the garden))))
```

```
(S (NP-SBJ There)
   (VP was
       (NP-PRD a sudden noise))
   .)
```

```
(SQ Are
    (NP-SBJ there)
    (NP-PRD (NP any flights)
            (VP arriving
                (PP-TMP after
                        (NP eleven a.m)))))
```

- With a relative clause.

```
(S (NP-SBJ There)
   (VP is
       (NP-PRD (NP nothing)
               (SBAR (WHNP-6 0)
                     (S (NP-SBJ *)
                        (VP to
                            (VP eat
                                (NP *T*-6)))))))
   .)
```

```
(S (NP-SBJ There)
   (VP has
       (VP been
           (NP-PRD (NP a lot)
                   (VP accomplished
                       (NP *)
                       (NP-TMP today))))))
```

- With a clausal complement or adjunct.

```
(S (NP-SBJ There)
   (VP is
       (NP-PRD no use)
       (S-ADV (NP-SBJ-1 *)
              (VP trying
                  (S (NP-SBJ *-1)
                     (VP to
                         (VP `` Explain ''
                             (PP-DIR to
                                     (NP a 2-year-old))))))))
```

```
(S (ADVP-TMP (NP A few weeks)
             ago)
   ,
   (NP-SBJ I)
   (VP read
       (PP-LOC in
               (NP the Bulletin))
       (SBAR that
```

```
(S (NP-SBJ-1 there)
   (VP were
       (S (NP-SBJ-224 *-1)
          (VP to
              (VP be
                  (VP given
                      (NP *-224)))))
       (NP-PRD Chinese classes)
       (PP-LOC in
               (NP Cranston)))))))
```

21 *Though*-Clefts

Though-clefts are considered a special case of topicalization, and so are bracketed as follows. (See section 4 [Null Elements] for more details.)

```
(S (SBAR-ADV (VP-TPC-2 Shout
                       (PP-CLR at
                               (NP Eichmann)))
             though
             (S (NP-SBJ he)
                (VP might
                    (VP *T*-2))))
   ,
   (NP-SBJ the Prosecutor)
   (VP could
       not
       (VP establish...)))


(S (SBAR-ADV (ADJP-PRD-TPC-1 Airless and dingy)
             though
             (S (NP-SBJ it)
                (VP was
                    (ADJP-PRD *T*-1))))
   ,
   (NP-SBJ the attic)
   (VP represented
       (NP luxury)))
```

22 Comparatives

This section describes comparative structures and related structures. Comparatives represent a complex and difficult problem, and the bracketing policy for comparatives was never finalized. As a result, variation in analysis is more prevalent for comparatives than for simpler constructions.

22.1 Basic tools for bracketing the *than/that/as*-phrase

The *than*, *that*, or *as* is bracketed as either a PP or an SBAR, and a certain amount of variation exists in the choice of PP or SBAR. SBAR is used when the rest of the *than/that/as*-phrase is a tensed sentence, or when it contains a subject. PP is in general used when the rest of the *than/that/as*-phrase is a single constituent. There is a tendency to use SBAR when the rest of the *than/that/as*-phrase is a VP or other predicate tagged -PRD (even if it is a single word) and when the rest of the phrase is dominated by FRAG, though PP may also be used.

The rest of the *than/that/as*-phrase after the *than*, *that* or *as* is most often bracketed simply with the bracket labels and function tags appropriate for the constituent. It may be dominated by FRAG, particularly if more than one constituent is involved or if the rest of the phrase is a VP or other predicate (but not an S).

The null element *?* is used to indicated missing constituents in the predicate of the *than/that/as*-phrase. (See section 4 [Null Elements] for a more complete description of the uses of *?*.) The null element *?* has the bracket label that the missing constituent would have if present (see section 22.5).

Throughout this section on comparatives, alternate bracketings are shown when they seem particularly likely or common (i.e., not all of the possible variants are shown for each example).

A schematic for possible bracketings follows:

1. PP or SBAR for the *than*, *that* or *as*.

   ```
   (PP than/as
       (xP rest of phrase))
   ```

   OR:

   ```
   (SBAR than/that/as
         (S rest of phrase))
   ```

2. For the rest of the phrase, after the *than*, *that* or *as*:

   - Appropriate bracket label and function tags.

     ```
     (PP than
         (NP ice cream))
     ```

     OR:

     ```
     (SBAR as
           (S (NP-SBJ I)
              (VP was)))
     ```

   - FRAG may dominate.

     ```
     (PP as
         (FRAG (ADJP possible)))
     ```

OR:

```
(SBAR as
      (FRAG (ADJP-PRD possible)))
```

- With predicates, the null element *?* may be used for missing parts of the predicate.

```
(SBAR as
      (S (NP-SBJ I)
         (VP was
             (ADJP-PRD *?*))))
```

22.2 Adjunction

The DEFAULT for most comparative constructions is to Chomsky-adjoin the *than/that/as*-phrase to the comparative phrase (ADJP, ADVP, NP). If the comparative phrase is an NP modifier, the *than/that/as*-phrase is adjoined to the entire NP rather than to the modifier.

Examples of several common constructions follow:

1. *AS*

   AS — AS

```
(ADJP (ADJP as mysterious)
      (PP as
          (NP (NP a tiny hole)
              (PP-LOC in
                      (NP my skin)))))
```

   AS MUCH — AS —

```
(S (NP-SBJ it)
   (VP might not
       (VP mean
           (NP (NP as much)
               (SBAR as
                     (S (NP-SBJ it)
                        (VP means
                            (PP to
                                (NP us)))))))))
```

   NOT SO MUCH — AS —

```
(NP (NP not
        (ADVP so much)
        a search)
    (PP for
        (NP truth))
    (PP as
        (PP for
            (NP certainty))))
```

AS [ADJP] A [NP/SBAR] AS

```
(NP (NP (ADJP as good) a year)
    (PP as
        (NP 1989)))
```

2. *THAN*

—ER THAN

```
(ADJP (ADJP friendlier)
      (PP than
          (NP (NP a dog)
              (PP on
                  (NP a picnic)))))
```

MORE THAN

```
(ADJP (ADJP more interesting)
      (SBAR than
            (S (NP-SBJ I)
               (VP thought))))
```

OR:

```
(ADJP (ADJP more interesting)
      (SBAR than
            (S (NP-SBJ I)
               (VP thought
                   (S *?*)))))
```

3. *THAT*

SO — THAT

```
(ADJP (ADJP so friendly)
      (SBAR that...))

(ADVP (ADVP so slowly)
      (SBAR that...))
```

SO — A — THAT

```
(NP (NP so playful a kitten)
    (SBAR that...))
```

SUCH A — THAT

```
(NP (NP such a playful kitten)
    (SBAR that...))
```

22.3 Items intervening between the comparative phrase and the *than/that/as*-phrase

22.3.1 Simple adjunction

The *than/that/as*-phrase is adjoined as usual to the comparative phrase if the intervening item is another modifier of the same comparative phrase (i.e., not attached at a higher level).

```
(S (NP-SBJ There)
   (VP 's
       (NP-PRD (NP more reading and instruction)
               (SBAR (WHNP-1 0)
                     (S (NP-SBJ-2 *T*-1)
                        (VP to
                            (VP be
                                (VP heard
                                    (NP *-2)
                                    (PP-LOC on
                                            (NP discs)))))))
               (PP than
                   (ADVP-TMP ever before))))
   .)
```

22.3.2 *ICH*-attachment

If the intervening item is attached at a higher level, the *than/that/as*-phrase is *ICH*-attached (and the *ICH* null element adjoined) to the comparative phrase.

```
(S (NP-SBJ it)
   (VP might not
       (VP mean
           (NP (NP as much)
               (SBAR *ICH*-3))
           (PP to
               (NP German banking))
           (SBAR-3 as
                   (S (NP-SBJ it)
                      (VP means
                          (PP to
                              (NP us)))))))
   . '')

( (S (NP-SBJ (NP More industrial acreage)
             (PP *ICH*-1))
     (VP lies
         (ADVP-CLR vacant)
         (PP-LOC in
                 (NP St. Clair county))
         (PP-1 than
               (PP-LOC in
                       (NP (NP any other jurisdiction)
                           (PP-LOC in
                                   (NP the St. Louis area))))))
```

```
.))
```

22.4 *Than/as*-phrase containing only one constituent

In most cases, when the *than/as*-phrase contains only one constituent, the *than* or *as* is bracketed as a PP with the single constituent as its complement.

However, when the single constituent is a predicate (i.e., a VP or -PRD), the *than* or *as* is often bracketed as an SBAR. The predicate may be immediately dominated by FRAG or S with a null * subject.

1. with NP

```
(S (NP-SBJ I)
   (VP like
       (NP cake)
       (ADVP (ADVP more)
             (PP than
                 (NP ice cream)))))
```

2. with PP

```
( (S (NP-TMP Last year)
     ,
     (NP-SBJ the average broker)
     (VP earned
         (NP (NP $ 71,309 *U*)
             ,
             (ADJP (ADJP (NP 24 %)
                         lower)
                   (PP than
                       (PP-TMP in
                               (NP 1987))))))
     .))
```

OR:

```
(ADJP (ADJP (NP 24 %) lower)
      (PP than
          (FRAG (PP-TMP in
                        (NP 1987)))))
```

3. with VP

```
( (S (NP-SBJ visitors)
     (VP have
         (NP (NP more)
             (SBAR (WHNP-1 0)
                   (S (NP-SBJ *)
                      (VP to
                          (VP do
                              (NP *T*-1)))))
```

```
               (PP than
                   (VP ski))))
       .))
```

OR:

```
(NP (NP more)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP do
                     (NP *T*-1)))))
    (SBAR than
          (S (NP-SBJ *)
             (VP ski))))
```

OR:

```
(NP (NP more)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP do
                     (NP *T*-1)))))
    (SBAR than
          (FRAG (VP ski))))
```

OR:

```
(NP (NP more)
    (SBAR (WHNP-1 0)
          (S (NP-SBJ *)
             (VP to
                 (VP do
                     (NP *T*-1)))))
    (PP than
        (FRAG (VP ski))))
```

4. with -PRD

```
( (S ``
     (NP-SBJ (NP It)
             (S *EXP*-2))
     (VP 's
         (ADJP-PRD (ADJP easier)
                   (SBAR *ICH*-1))
         (S-2 (NP-SBJ *)
              (VP to
                  (VP get
                      (ADJP-PRD worse))))
         (SBAR-1 than
                 (FRAG (ADJP-PRD better))))
     . ''))
```

5. with *expected*

```
(S (NP-SBJ (NP The total)
           (PP of
               (NP (NP 18 deaths)
                   (PP from
                       (NP (NP malignant mesothelioma)
                           ,
                           (NP lung cancer)
                           and
                           (NP asbestosis))))))
   (VP was
       (ADJP-PRD (ADJP far higher)
                 (SBAR than
                       (S (NP-SBJ *)
                          (VP expected)))))
   .)
```

OR:

```
(ADJP-PRD (ADJP far higher)
          (SBAR than
                (S (NP-SBJ-1 *)
                   (VP expected
                       (NP *-1)))))
```

OR:

```
(SBAR than
      (FRAG (VP expected)))
```

NOTE: if *expected* occurs in a fleshed-out sentence, a *?* is likely to be used, as in:

```
(NP (NP fiscal fourthquarter earnings)
    (SBAR (WHNP-1 that)
          (S (NP-SBJ *T*-1)
             (VP were
                 (ADJP-PRD (ADJP better)
                           (SBAR than
                                 (S (NP-SBJ analysts)
                                    (VP had
                                        (VP expected
                                            (S *?*))))))))))
```

6. with *possible*, *usual*, etc.

Usually done as a PP with an ADJP complement:

```
(NP (NP more)
    (PP than
        (ADJP usual)))
```

```
(NP (NP as many
        (ADJP solidly minority)
        districts)
    (PP as
        (ADJP possible)))
```

Rarely done as SBAR:

```
(NP (NP as many
        (ADJP solidly minority)
        districts)
    (SBAR as
          (FRAG (ADJP-PRD possible))))
```

22.5 More complicated *than/as*-phrases — use of *?*

When the *than/as*-phrase contains both a subject and a portion of a predicate, these constituents form the basis of an S, and the missing elements (i.e., the elements which are interpreted but not realized) are often represented by *?*. (See section 4 [Null Elements] for more details on *?*.)

In the following list, the likelihood of there being a *?* goes from greatest to least.

1. subject / copular verb / missing predicate (most likely use for *?*)

   In this example, the missing predicate is a PP, assumed to be something like *of military value*.

```
( (S (NP-SBJ Laos)
     (VP is
         (PP-PRD of
                 (NP (NP (ADJP (ADVP no more)
                               purely military)
                         value)
                     (SBAR *ICH*-2)))
         (PP to
             (NP (NP Moscow)
                 (NP itself)))
         (SBAR-2 than
                 (S (NP-SBJ it)
                    (VP is
                        (PP-PRD *?*)
                        (PP to
                            (NP Washington))))))
     .))
```

2. subject / other main verb / missing direct object

```
( (S (NP-SBJ the Controller)
     (VP will
         (VP have
             (NP (NP the opportunity)
                 (PP for
                     (NP (NP greater usefulness)
```

```
                              (PP to
                                  (NP good government))
                              (SBAR than
                                    (S (NP-SBJ he)
                                       (VP has
                                           (NP *?*)
                                           (ADVP-TMP now))))))))
          .))
```

3. subject / auxiliary / missing main verb

```
(S (NP-SBJ The submission)
   (VP would
       (VP place
           (NP the issues)
           (PP-LOC-CLR before
                       (NP the court))
           (ADVP (ADVP more readily)
                 (SBAR than
                       (SINV would
                             (NP-SBJ (NP discussion)
                                     (PP in
                                         (NP the abstract)))
                             (VP *?*))))))
   .)
```

4. subject / main verb / missing clausal complement

```
(S (NP-SBJ the steel strike)
   (VP lasted
       (ADVP-TMP (ADVP much longer)
                 (SBAR than
                       (S (NP-SBJ he)
                          (VP anticipated
                              (SBAR 0
                                    (S *?*)))))))
   .)
```

OR:

```
(S (NP-SBJ the steel strike)
   (VP lasted
       (ADVP-TMP (ADVP much longer)
                 (SBAR than
                       (S (NP-SBJ he)
                          (VP anticipated
                              (SBAR *?*))))))
   .)
```

5. subject / auxiliary / auxiliary replaces main verb

```
(S (NP-SBJ Bill)
   (VP ate
       (NP (NP more hotdogs)
           (SBAR than
                 (S (NP-SBJ Mary)
                    (VP did
                        (VP *?*
                            (NP-TMP yesterday)))))))
   .)
```

```
( (S (NP-SBJ Bill)
     (VP eats
         (NP (NP more hotdogs)
             (SBAR than
                   (S (NP-SBJ Mary)
                      (VP does
                          (VP *?*))))))
     .))
```

OR:

```
( (S (NP-SBJ Bill)
     (VP eats
         (NP (NP more hotdogs)
             (SBAR than
                   (S (NP-SBJ Mary)
                      (VP does)))))
     .))
```

22.6 Superlative + relative clause

Superlatives with relative clauses are bracketed using the standard bracketing for an NP with a relative clause modifying it. There is no comparative structure shown.

```
(S (NP-SBJ He)
   (VP was
       (ADVP altogether)
       (NP-PRD (NP the (ADJP most combustible looking) man)
               (SBAR (WHNP-1 0)
                     (S (NP-SBJ I)
                        (ADVP-TMP ever)
                        (VP saw
                            (NP *T*-1)))))))
```

23 "Financialspeak" conventions

This section covers some of the constructions that are specific to texts about financial happenings (hereafter, referred to as "Financialspeak"). Note that some of the treatments described below are not found in any context outside Financialspeak.

23.1 Salient features of Financialspeak text

Annotators determine intuitively whether a particular set of tokens is Financialspeak. Text that annotators consider to be Financialspeak tends to have one or more of the following characteristics.

- It contains verbs like *rise, grow, increase, decrease, drop, fall, jump, close, finish*, etc.
- the entire file is a list of rising and falling stock or bond prices
- rising and falling monies
- repetitious sentence structures, centering around the above verbs
- it's talking about financial stuff, but you don't really know what's going on
- there's no sensible way to interpret the sentence other than as Financialspeak.

Though the Treebank has no precise way of delimiting Financialspeak, there is a reasonable degree of agreement among annotators about when they are bracketing Financialspeak.

23.2 Bracketing conventions

23.2.1 Bracketing of *up/down*-phrases in VP

1. *Up/down*-phrases are bracketed ADVP-CLR when they immediately follow the Financialspeak verbs listed above (particularly *closed, settled* and *finished*) and when there are no other constituents. The quantificational NP (i.e., *5 points*) is attached as a complement of *up* or *down*.

```
(S (NP-SBJ IBM)
   (VP closed
       (ADVP-CLR up
                 (NP 5 points))))


(S (NP-SBJ IBM)
   (VP finished
       (ADVP-CLR down
                 (NP 5 points))))
```

   After copular verbs, *up/down*-phrases are labeled -PRD.

```
(S (NP-SBJ IBM)
   (VP was
       (ADVP-PRD up
                 (NP 5 points))))
```

Non-*up/down*-phrases which are interpreted as describing the subject in some way, and which follow non-copular verbs, may be interpreted as secondary predication, as follows.

```
(S (NP-SBJ-1 IBM)
   (VP finished
       (S-ADV (NP-SBJ *-1)
              (ADJP-PRD unchanged)))
   .)
```

2. If the VP contains an *up/down*-phrase followed by another modifier, there is no defined policy dictating which item(s) receive -CLR. The corpus may contain the following bracketings.

```
(S (NP-SBJ Copper)
   (VP finished
       (ADVP-CLR down
                 (NP 4.5 cents))
       ,
       (PP-CLR at
               (NP (NP $ 1.2345 *U*)
                   (NP-ADV a pound))))
   .)

(S (NP-SBJ Copper)
   (VP finished
       (ADVP-CLR down
                 (NP 4.5 cents))
       ,
       (PP at
           (NP (NP $ 1.2345 *U*)
               (NP-ADV a pound))))
   .)
```

Note that constructions like *4.5 cents lower* are treated the same as *up/down*-phrases.

```
(S (NP-SBJ Copper)
   (VP finished
       (ADVP-CLR (NP 4.5 cents)
                 lower)
       ,
       (PP-CLR at
               (NP (NP $ 1.2345 *U*)
                   (NP-ADV a pound))))
   .)
```

3. If an item intervenes between an *up/down*-phrase and the verb, and the *up/down*-phrase is a child of VP (see the next section for "Attachment"), the *up/down*-phrase does not receive -CLR.

```
(S (NP-SBJ-1 (NP Volume)
             (PP-LOC on
                     (NP the first section)))
   (VP was
       (VP estimated
           (NP *-1)
           (PP-CLR at
```

```
                    (NP (QP 1 billion) shares))
                ,
                (ADVP up
                      (PP from
                          (NP (QP 914 million))
                          (NP-TMP Tuesday))))))
```

4. Attachment: child of VP versus NP adjunct

   When items intervene between an *up/down*-phrase and the verb, the *up/down*-phrase is attached at VP level if it is possible to say: "[SBJ] [to be] [ADVP]". As mentioned in the preceding section, the *up/down* phrase does not receive -CLR.

```
(S (NP-SBJ Sales)
   (VP were
       (NP-PRD (QP $ 1.25 billion) *U*)
       ,
       (ADVP down
             (PP from
                 (NP (QP $ 1.36 billion) *U*)
                 (PP-TMP in
                         (NP the 1988 quarter))))))

(S (NP-SBJ IBM)
   (VP rose
       (PP-DIR to
               (NP 101))
       ,
       (ADVP up
             (NP 3 %))))

(S (NP-SBJ The Financial Times 100-share index)
   (VP finished
       (PP-CLR at
               (NP 2161.9))
       ,
       (ADVP up
             (NP 12.6 points))))
```

   Otherwise, (i.e., if one can't say "[SBJ] [to be] [ADVP]"), the *up/down*-phrase is attached to the NP. (*The U.S. is up 2 billion from a year ago.)

```
(S (NP-SBJ The U.S.)
   (VP imported
       (NP (NP (QP 6 billion) barrels)
           (PP of
               (NP oil))
           (ADVP *ICH*-1))
       (NP-TMP this year)
       ,
       (ADVP-1 up
               (NP (QP 2 billion))
               (PP from
                   (ADVP-TMP (NP a year)
                             ago)))))
```

23.2.2 -CLR on PPs associated with *sold, bought, estimate, priced*

```
(S (NP-SBJ-1 The Water Works)
   (VP was
       (VP sold
           (NP *-1)
           (PP-CLR for
                   (NP $500)))))

(S (NP-SBJ-1 500 shares)
   (VP were
       (VP bought
           (NP *-1)
           (PP-CLR from
                   (NP the unsuspecting old lady)))))

(S (NP-SBJ (NP PAFA's)
           stock)
   (VP sold
       (PP-CLR at
               (NP (NP $ 17 *U*)
                   (NP-ADV a share)))))

(S (NP-SBJ-1 (NP convertible debentures)
             (PP of
                 (NP AB&C)))
   (VP were
       (VP priced
           (NP-2  *-1)
           (PP-CLR at
                   (NP $ 400 *U*))
           (S-CLR (NP-SBJ *-2)
                  (VP to
                      (VP yield
                          (NP 8%)))))))
```

23.2.3 PP-DIR and double complements

1. The prepositions *to* and *from* are labeled PP-DIR when they are complements of a Financialspeak verb.

2. Financialspeak *to* and *from* are analyzed as optionally taking two complements: a range and a time. The time will always be tagged -TMP, and the range will have no dash tag. The possibilities are

   (a) *to/from* RANGE, TIME

   (b) *to/from* RANGE

   (c) *to/from* TIME

3. Both RANGE and TIME will be bracketed as complements of the preposition (i.e., attached at the same level, as children of the PP).

4. Examples:

- RANGE

```
(S (NP-SBJ IBM)
   (VP rose
       (NP-EXT 3 %)
       (PP-DIR to
               (NP (QP 101 1/2)))))
```

- TIME

```
(S (NP-SBJ (NP The
               (ADJP West German)
               machinery and plant equipment industry 's)
           orders)
   (VP rose
       (NP-EXT an inflation-adjusted 1 %)
       (PP-TMP in
               (NP September))
       (PP-DIR from
               (NP-TMP a year earlier))))
```

- RANGE, TIME

```
( (S (NP-SBJ Annual inflation)
     (VP rose
         (PP-DIR to
                 (NP 3.64 %)
                 (PP-TMP in
                         (NP October)))
         (PP-DIR from
                 (NP 3.55 %)
                 (PP-TMP in
                         (NP September))))
     .))
```

5. Miscellaneous note:

   Only in Financialspeak, and only when the "mother PP" is a PP-DIR, can temporal modifiers be put inside a PP as in this example:

```
(S (NP-SBJ unconsolidated pretax profit)
   (VP increased
       (NP-EXT 70 %)
       (PP-DIR to
               (NP (QP 12.12 billion) yen))
       (PP-TMP in
               (NP (NP the first half)
                   (VP ended
                       (NP-TMP Sept. 30))))
       ,
       (PP-DIR from
               (NP (QP 7.12 billion) yen)
               (ADVP-TMP (NP a year)
                         ago))))
```

   Otherwise, -TMP things go at S or VP level:

```
(S (PP-LOC In
           (NP active trading))
   (PP-TMP on
           (NP Tuesday))
   ,
   (NP-SBJ IBM)
   (VP vanished))


(S (NP-SBJ Bats)
   (VP invaded
       (NP my apartment)
       (PP-TMP at
               (NP midnight))
       (NP-TMP Friday)))
```

23.3 Financialspeak bracketing conventions covered elsewhere in this manual:

For information about QPs, see section 11 [Modification of NP].

For information about -EXT, see section 2 [Notation].

24 Numbered Lists

The bracketings described below for lists that include list item markers also generalize somewhat to similarly long lists that lack item markers.

1. LST

   Letters and numerals which identify items in a list, and their surrounding punctuation, are labeled LST:

```
(LST -LRB-
     1
     -RRB-)
```

2. List item markers include:

```
      1, 2, 3                 a, b, c
      i, ii, iii              one, two, three
```

3. The list item marker is made a child of the constituent it precedes.

4. When the enumerated items are listed in one sentence, they are conjoined:

```
( (S (NP-SBJ-1 It)
     (VP was
         (VP used
             (NP *-1)
             (S-CLR (NP-SBJ *)
                    (VP (VP (LST -LRB-
                                 1
                                 -RRB-)
                            to
                            (VP investigate
                                (NP wave behavior)))
                        ,
                        (VP (LST -LRB-
                                 2
                                 -RRB-)
                            to
                            (VP estimate
                                (NP the wave energy)))
                        ,
                        and
                        (VP (LST -LRB-
                                 3
                                 -RRB-)
                            forecast
                            (NP coastal changes))))))
     .))
```

5. When the enumerated items occur in separate sentences (i.e. each list item ends with a period or some other kind of final punctuation), treat the colon as final punctuation and place each list item in its own set of empty outer parentheses:

```
( (S (NP-SBJ The aged care plan)
     (VP carries
         (NP these benefits)
         (PP for
             (NP (NP persons)
                 (PP over
                     (NP 65)))))
     :))
( (NP (LST 1)
      (NP Full payment)
      (PP of
          (NP (NP hospital bills)
              (PP for
                  (NP (NP stays)
                      (NP (QP up to 90) days)))))
      .))
( (NP (LST 2)
      (NP Full payment)
      (PP of
          (NP nursing home bills))
      (PP-TMP for
              (NP (NP (QP up to 180) days)
                  (PP-TMP following
                          (NP (NP discharge)
                              (PP from
                                  (NP a hospital))))))
      .))
( (NP (LST 3)
      (NP Hospital outpatient clinic diagnostic service)
      (PP for
          (NP (NP all costs)
              (PP in
                  (NP (NP excess)
                      (PP of
                          (NP (NP $ 20)
                              (NP-ADV a patient)))))))
      .))
```

6. Lists in apposition:

```
( (PP-PRP for
          (NP (NP several reasons)
              :
              (NP (NP (LST 1)
                      (NP (NP Broglio 's)
                          (NX (NX 4-0 won-lost record)
                              and
                              (NX 1.24 earned-run mark)))
                      (PP against
```

```
                  (NP Pittsburgh))
              (ADVP-TMP (NP a year)
                        ago))
          ;
          (NP (LST 2)
              The desire
              (S (NP-SBJ *)
                 (VP to
                     (VP give
                         (NP Broglio)
                         (NP (NP (ADJP as many)
                                 starts)
                             (PP as
                                 (ADJP possible)))))))
          ;
          (NP (LST 3)
              (NP (NP The Redbirds ')
                  disheartening 11-7 collapse)
              (PP against
                  (NP the Phillies))
              (NP-TMP Sunday))))))
```

The below contains an example of a non-numbered list. Treatment is similar to that for numbered lists; the list items are adjoined if not separated by final punctuation, and the adjoined list is in turn adjoined to the introducing phrase.

In this case, the introducing phrase is not adjacent to the list, so the list (but *not* the separating colon) is *ICH*-attached. Finally, the entire list is grouped under NP because -TTL implies -NOM and multiples -NOMs are adjoined under NP.

```
( (S (NP-SBJ-1 (NP 3 books)
               (NP *ICH*-2))
     (VP are
         (VP recommended
             (NP *-1)
             (PP-MNR with
                     (NP gusto))
             :
             (NP-2 (NP-TTL Crime and Punishment)
                   ,
                   (S-TTL (NP-SBJ One)
                          (VP Flew
                              (PP Over
                                  (NP (NP the Cuckoo 's)
                                      Nest))))
                   and
                   (NP-TTL (NP The House)
                           (PP-LOC at
                                   (NP Pooh Corner))))))
     .))
```

25 Correlative *the*-Clauses

Treatment of proportional clauses introduced by the fronted correlative *the...the* is based upon the following schematic bracketing:

```
(S (SBAR-ADV (X the sooner)
             (S our vans hit the road each morning))
   ,
   (X the easier)
   it is for us to fulfill that obligation
   .)
```

Annotators were to try to approximate the above bracketing as best they could for *the...the* constructions encountered in the WSJ corpus. *The...the* policy was not a high priority for the Treebank, due to the rarity and irregular nature of these constructions, so a more specific policy does not exist.

The following examples show how some sentences from the corpus **might** have been bracketed:

```
( (S (SBAR-ADV (X The sooner)
               (S (NP-SBJ our vans)
                  (VP hit
                      (NP the road)
                      (NP-TMP each morning))))
     ,
     (X the easier)
     (NP-SBJ (NP it)
             (SBAR *EXP*-1))
     (VP is
         (ADJP-PRD *?*)
         (SBAR-1 for
                 (S (NP-SBJ us)
                    (VP to
                        (VP fulfill
                            (NP that obligation))))))
     .))


(S (SBAR-ADV (X the more)
             (S (NP-SBJ a scandal)
                (VP has
                    (S (NP-SBJ *)
                       (VP to
                           (VP do
                               (PP-CLR with
                                       (NP (NP (NP a congressman 's)
                                               duties)
                                           (PP as
                                               (NP a congressman))))))))))
   ,
   (X the
      (ADJP less likely))
   (NP-SBJ it)
   (VP is
       (S (NP-SBJ *)
```

```
          (VP to
              (VP catch
                  (NP (NP the fancy)
                      (PP of
                          (NP a network)))))))

(S (SBAR-ADV (X the more)
             (S (NP-SBJ he)
                (VP muzzles
                    (NP his colleagues))))
   ,
   (X the more)
   (NP-SBJ leaks)
   (VP will
       (VP pop
           (PRT up)
           (PP-LOC all around
                   (NP Washington)))))

( (S ``
     (SBAR-ADV (X The less)
               (S (NP-SBJ they)
                  (VP know
                      (NP *?*))))
     ,
     (X the easier)
     (NP-SBJ it)
     (VP is
         (PP for
             (NP us)))
     .))

(S (SBAR-ADV (X The more)
             (S (NP-SBJ we)
                (VP think
                    (PP-CLR about
                            (NP it)))))
   ,
   (X the more)
   (NP-SBJ we)
   (VP suspect
       (SBAR 0
             (S (NP-SBJ Mr. Brady)
                (VP does
                    (ADVP indeed)
                    (VP have
                        (NP enough power)
                        (SBAR-LOC (WHADVP-1 where)
                                  (S (NP-SBJ he)
                                     (ADVP-TMP already)
                                     (VP is
                                         (ADVP-LOC-PRD *T*-1))))))))))
```

```
   .)

(S (NP-SBJ A trader)
   (VP said
       (SBAR that
             (X (SBAR-ADV (X the more)
                          (S (NP-SBJ an issue)
                             (VP gained
                                 (NP *?*)
                                 (ADVP-TMP recently))))
                ,
                (X the sharper)
                (NP (NP the loss)
                    (VP sustained
                        (NP *)
                        (NP-TMP Wednesday))))))
   .)

( (S (SBAR-ADV (X The more factories and robots)
               (S (NP-SBJ Japanese manufacturers)
                  (VP add
                      (NP *?*))))
     ,
     (S (X the more)
        (NP-SBJ-1 they)
        (VP will
            (VP be
                (ADJP-PRD able
                          (S (NP-SBJ *-1)
                             (VP to
                                 (VP export
                                     (NP *?*))))))))
     , and
     (S (X the less)
        (NP-SBJ-2 their domestic customers)
        (VP will
            (VP need
                (S (NP-SBJ *-2)
                   (VP to
                       (VP import
                           (NP *?*)))))))
     .))
```

26 Orphans

This section includes miscellaneous constructions that have not found a home under other headings, and other oddities.

26.1 List of miscellaneous phrases

all about:

```
(S (PP-TMP from
           (NP childhood))
   (NP-SBJ he)
   (VP had
       (VP known
           (PP-CLR all about
                   (NP knives)))))
```

as of:

```
(PP-TMP as
        (PP of
            (NP the first)))
```

come (as in *"come spring cleaning..."*): PP

```
(S (NP-SBJ You)
   (VP can
       (VP hope
           (PP against
               (NP hope))
           (SBAR that
                 (S (PP-TMP come
                            (NP spring cleaning))
                    ,
                    (NP-SBJ your fair-weather friends)
                    (VP will))))))
```

effective:

The word *effective* occasionally introduces an adverbial with a time complement, as in *"The chairman is resigning effective Monday."* These are fairly rare, so no uniform treatment exists.

Some are analyzed as "floating participials":

```
(S (NP-SBJ-2 he)
   (VP is
       (VP quitting
           (S-ADV (NP-SBJ *)
                  (ADJP-PRD effective
                            (NP-TMP Dec. 31))))))
```

while others are simply ADJP, often with a -ADV or -TMP tag:

```
(S (NP-SBJ-34 Columbia Pictures Entertainment Inc.)
   (VP was
       (VP dropped
           (NP *-34)
           ,
           (ADJP-ADV effective
                     (NP-TMP today))
           ,
           (PP-DIR from
                   (NP the recreational products group)))))
```

half (see also the section on "Multipliers" in section 11 [Modification of NP]):

"...using half whole wheat and half white flour"

```
(S (NP-SBJ *)
   (VP using
       (NP (NP half whole wheat
               (NX *RNR*-1))
           and
           (NP half white
               (NX *RNR*-1))
           (NX-1 flour))))
```

no doubt:

1. as adverb

```
(S (NP-SBJ He)
   (NP-ADV no doubt)
   (VP will
       (VP go)))
```

```
(S (NP-SBJ He)
   (VP 'll
       (NP-ADV no doubt)
       (VP go)))
```

2. as noun

```
(S (NP-SBJ There)
   (VP 's
       (NP-PRD no doubt
               (SBAR that
                     (S (NP-SBJ he)
                        (VP 'll
                            (VP go)))))))
```

no matter:

```
(S (PP-LOC In
           (NP the stands))
   (NP-SBJ he)
   (VP is
       (ADJP-PRD lonely and lost)
       ,
       (ADVP no matter
             (SBAR (WHNP-1 how many)
                   (S (NP *T*-1)
                      (VP are
                          (PP-LOC-PRD about
                                      (NP him))))))))
```

percent:

Percent is simply a flat NP, whether or not it is written with a space:

```
(NP 15 percent)
```

```
(NP 15 per cent)
```

regardless of:

```
(ADVP regardless
      (PP of ...))
```

though:

1. clefted: see section 21 [*Though*-clefts] and section 4 [Null Elements] for more on *though*-clefts.
2. fronted: SBAR-ADV

   ```
   (SBAR-ADV Though (FRAG (ADJP limited)))
   ```

3. not fronted: conjunction

   ```
   (NP (NP the
           (ADJP well-defined
                 though
                 limited)
           range)
       (PP of
           (NP motifs)))
   ```

using:

```
(VP estimated
    (S-MNR (NP *)
           (VP using)))
```

worth:

1. with complement: ADJP

   Note that some instances of this use of *worth* are labeled PP-PRD, as in (b); however the use of ADJP-PRD, as in (a), predominates.

   (a)
```
(S (PP With
       (NP (NP respect)
           (PP to
               (NP this view))))
   ,
   (NP-SBJ two points)
   (VP are
       (ADJP-PRD worth
                 (S (NP-SBJ *)
                    (VP making)))))


(S (NP-SBJ (NP the results)
           ,
           (ADJP however general)
           ,)
   (VP are
       (ADJP-PRD worth
                 (NP the search))))
```

   (b)
```
(S (PP With
       (NP (NP respect)
           (PP to
               (NP this view))))
   ,
   (NP-SBJ two points)
   (VP are
       (PP-PRD worth
               (S-NOM (NP-SBJ *)
                      (VP making)))))


(S (NP-SBJ (NP the results)
           ,
           (ADJP however general)
           ,)
   (VP are
       (PP-PRD worth
               (NP the search))))
```

2. *dollars worth*: NP

   There is considerable variation, but here is a common way of analyzing expressions like *five dollars worth*:

```
(VP issue
    (NP (NP (ADJP (QP some $ 3 million to $ 4 million) *U*)
            worth)
```

```
        (PP of
            (NP Rural Roads Authority bonds))))
```

26.2 Flat multi-word ADVPs and PPs

1. The following are annotated as flat ADVPs:

   all but

   at all

   at best

   at least

   at most

   more than (...doubled)

   Unfortunately, the POS tagging for these is more often (though not always) compositional: *at/IN all/DT*. See the *POS guidelines* [Santorini 1990] for more information.

2. The following are annotated as flat PPs:

   because of

   instead of (may also be CONJP)

   rather than (may also be CONJP)

   such as

   ```
   (NP (NP desserts)
       (PP such as
           (NP (NP ice cream)
               and
               (NP brownies))))
   ```

26.3 Foreign words

There is no special bracket tag for foreign words. They are simply bracketed according to the annotator's interpretation of their syntactic function. If the annotator is unable to determine its syntactic function, the phrase is labeled X. The internal structure of foreign phrases is not annotated.

```
( (S (NP-SBJ My favorite)
     (VP is
         (NP-PRD (NP pie)
                 (PP a la mode)))
     .))

( (S (NP-SBJ The new movie)
     (VP is
         (NP-PRD a tour de force))
     .))
```

```
(NP (NP ballooning)
    (PP-TMP at
            (NP (NP the (ADJP de rigueur) hour)
                (PP of
                    (NP 6 a.m.)))))
```

26.4 Negation

The negative element *not* is always left unlabeled and is attached in accordance with the policy governing the attachment of all adverbials. See reduced relatives sections in section 13 [Gerunds and Participles] and in section 8 [Shared Complements and Modifiers] for more information about negation in those cases.

```
( (S (NP-SBJ I)
     (VP do
         not
         (VP understand))
     .))
```

```
( (S (NP-SBJ She)
     (VP is
         not
         (NP-PRD a certified teacher))
     .))
```

```
( (S (NP-SBJ She)
     (VP is
         not
         (VP listening))
     .))
```

```
( (S (NP-SBJ I)
     (VP am
         (VP (VP going
                 (PP-NOM to (NP that other restaurant)))
             and
             (VP not
                 eating
                 (NP overpriced , overcooked broccoli stalks)
                 (ADVP-TMP again))))
     .))
```

```
(S (NP-SBJ I)
   (VP am
       (VP (VP going (PP to (NP that other restaurant)))
           (CONJP but not)
           (VP ordering (NP their broccoli))))
   .)
```

References

[Marcus et al. 1994] Marcus, M., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K. and Schasberger, B., 1994. The Penn Treebank: Annotating Predicate Argument Structure, in *Proceedings of the Human Language Technology Workshop, March 1994*, Morgan Kaufmann Publishers Inc., San Francisco, CA.

[Marcus et al. 1993] Marcus, M., Santorini, B., Marcinkiewicz, M.A., 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, Vol 19.

[Quirk et al. 1985] Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J., 1985. *A comprehensive grammar of the English language*, Longman, London.

[Santorini 1990] Santorini, B., 1990. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision)*, Tech Report MS-CIS-90-47, Linc Lab 178, Department of Computer and Information Science, University of Pennsylvania, Philadelphia.

Index