Supplementary Guidelines for ETTB 2.0

Justin Mott, Colin Warner, Ann Bies; Ann Taylor
Linguistic Data Consortium; University of York
{jmott|colinw|bies}@ldc.upenn.edu; at9a@york.ac.uk

April 6th, 2009

Contents

1 Tokenization

1.1 Overview

All strings separated by white space are treated as separate tokens. Also, no token can contain white space. The following contractions and related items are split into separate tokens. So, *women's*, *would've*, *cannot*, *he'll* are tokenized as `women 's`, `would 've`, `can not`, `he 'll`.

's
've
're
'll
'd
n't
can not
gon na
got ta
lem me
more 'n
't is
't was
wan na
wha dd ya

3

wha t cha

1.2 Hyphenated Words

Most hyphenated words are split into multiple tokens. For example, *elementary-school-age* is now treated as five tokens, viz.- `elementary(1) -(2) school(3) -(4) age(5)`. Hyphenated interjections and affixes in the following list are not split into multiple tokens. For example, *uh-oh* and *e-mail* are both single tokens: `uh-oh`, `e-mail`.

e-
a-
u-
x-
anti-
agro-
be-
bi-
bio-
co-
counter-
cross-
cyber-
de-
eco-
-esque
ex-
extra-
-fest
-fold
-gate
inter-
intra-
-itis
-less
macro-
mega-
micro-
mid-
mini-
mm-hm

4

mm-mm
-most
multi-
neo-
non-
o-kay
over-
pan-
para-
peri-
post-
pre-
pro-
pseudo-
quasi-
-rama
re-
semi-
sub-
super-
tri-
uh-huh
uh-oh
ultra-
un-
uni-
vice-
-wise

1.3   Punctuation

All other punctuation not described above triggers a break in tokenization, with the exceptions outlined below. Note that for present purposes, all non-alphanumeric characters are considered 'punctuation'.

1. Periods marking abbreviations.

    Mr. Dr. Ste. Ave. etc. e.g. A.D.

2. Punctuation in web addresses.

http://www.islamonline.net/Arabic/news/2004-12/05/images/pic05b.jpg
http://www.mofa.gov.sa/detail.asp?InNewsItemID=59090&InTemplateKey=print
rayhanenajib@menara.ma

3. Ellipses, when encoded as a string of periods. In addition, ellipses enclosed by round brackets are treated as single tokens in Sinorama.

...
(...)

4. Complex numerals.

2.45 20,000

5. Telephone numbers and postal codes.

tel : 02-2348-2192

6. Single quotation marks as parts of names.

the 80's
P'yongyang
'Assad

1.3.1 Punctuation in Webtext

Since the use of punctuation in webtext differs markedly from usage in carefully-edited text, it is handled slightly differently. The tokenization of punctuation in webtext is determined by whitespace boundaries. That is, all strings of punctuation (with the exception of quotation marks) without intervening whitespace are treated as a single token. So, a sequence such as *!!!!!!!!!!!????????* is treated as one token, rather than nineteen. Examples in include the following.

;)
:-)
..
?!
!!!!!!!!!!????????
****
_____

2 Part of Speech

2.1 Overview

This section summarizes annotation policy for POS-tagging developed in the course of annotation the English-Chinese Treebank (ECTB) and the English-Arabic Treebank (EATB), collectively referred to as the English Translation Treebank (or ETTB).

There are a total of 36 POS tags used. Definitions of most tags can be found in the original POS manual; tags unique to speech data (GW) can be found in the Switchboard manual; tags first introduced in Biomedical Treebanking (HYPH and AFX) can be found in BioMedical guidelines. Note also that the POS tag $, found in the original Penn Treebank, has been eliminated.

> **VB VBZ VBP VBD VBN VBG**
> **NN NNS NNP NNPS**
> **JJ JJR JJS**
> **RB RBR RBS**
> **PRP PRP$**
> **DT PDT**
> **IN**
> **RP**
> **POS**
> **CD**
> **EX**
> **TO**
> **, . : -LRB- -RRB- " "**
> **HYPH**
> **SYM**
> **LS**
> **AFX**
> **GW**
> **XX**
> **UH**
> **ADD NFP** (only used in webtext)

2.2 Nouns and Nominals

2.2.1 Nouns and Adjectives: NN* vs. JJ

Unambiguous adjectives are tagged JJ when they are the heads of NPs.

> the working poor/JJ
> only the lonely/JJ
> the wise/JJ
> in brief/JJ
> in particular/JJ
> at present/JJ
> in general/JJ
> and the like/JJ

For items that are ambiguous, the policy is as follows.

1. Heads of noun phrases are tagged NN.

   > the U.S. military/NN
   > a dangerous chemical/NN
   > a human/NN
   > gaudy material/NN
   > a wealthy executive/NN
   > the original/NN
   > research associate/NN
   > October 30th/NN

2. Modifiers are tagged JJ.

   > military/JJ strike
   > chemical/JJ reaction
   > human/JJ behavior
   > ceramic/JJ fibers
   > the chief/JJ task
   > a sovereign/JJ nation
   > the executive/JJ branch
   > associate/JJ professor
   > an original/JJ painting
   > grassroots/JJ organization
   > express/JJ mail
   > outside/JJ world
   > mini/JJ disc

30th/JJ time

2a. If the annotator has a **strong** intuition that the item is nominal rather than adjectival, it can be tagged NN.

a chemical/NN factory (i.e., a factory that produces chemicals)
a material/NN vendor (i.e., an organization that sells materials)
executive/NN privilege (i.e., a privilege possessed by executives)

Demonyms   It is useful to distinguish the following classes of demonyms:
Adjectival: Predominately ending in *-ese*. These cannot take plural morphology or head a noun phrase; they should always be tagged JJ. Examples include:*Chinese, Portuguese, Javanese*, etc.

*a Chinese
many Chinese/JJ
the Chinese/JJ people

Invariable: These have a single form for both nominal and adjectival uses. They can, in addition, take plural morphology and head a noun phrase. Examples include: *American, Thai, Pakistani, Arab*, etc. The tags of these are determined by position as outlined above.

an American/NNP
many Americans/NNPS
the American/JJ people

Variable: These have separate forms for adjectival and nominal usages. Examples include: *Swedish~Swede, Spanish~Spaniard, Croatian~Croat*, etc. The adjectival forms should be tagged as JJ, and the nominal forms as NNP.

a Swede/NNP (vs. *a Swedish)
many Swedes/NNPS (vs. *many Swedish)
the Swedish/JJ people (vs. *the Swede people)

The names of languages, however, are always tagged as proper nouns:

the French/JJ government
the French/NNP verb
World Englishes/NNPS

9

Adjectives in Proper Names: NNP vs. JJ   Expanding on the original POS guidelines for items like *United/NNP States/NNP,* the noun-adjective distinction is leveled in proper names. Adjectives which are part of the proper name are tagged NNP. To determine whether a capitalized adjective is part of the proper name it is useful for annotators to apply the "sign" test. That is, if the item in question would appear on a sign for the entity, then it is part of the name and tagged NNP. Otherwise, it is a descriptor of the entity and tagged JJ.

> American/NNP University in D.C.
> American/JJ Secretary of State Hillary Clinton
> Swedish/NNP Fish/NNPS candy
> Swedish/JJ Prime Minister
> Syrian/JJ Army/NNP team
> the Israeli/NNP Archaelogy/NNP department/NN
> Islamic/NNP Revolution

Cardinal Directions   For cardinal directions, forms ending in *-ern* are tagged as adjectives; unsuffixed forms are tagged as nouns.

> eastern/JJ Jerusalem
> east/NN Jerusalem
> southeastern/JJ Pennsylvania
> West/NNP Philly

2.2.2   Common vs. Proper Nouns: NN(S) or NNP(S)

In general, annotators rely on capitalization to determine whether a given item should be tagged NN or NNP. This is, however, not an ironclad rule since (parts of) some proper names are not capitalized and not all capitalized nouns are (parts of) proper names.

Capitalized acronyms which would not be capitalized if written out fully (such as *PR* for *public relations*) are tagged NN rather than NNP. Similarly, if the capitalization seems to be a quirk of the editor/translator, it should be tagged NN. Capitalized letters used as designations are tagged NN.

> RMB/NN
> GDP/NN
> MP3/NN
> CD/NN
> CD-ROM/NN
> PR/NN firm

works in R&D/NN
PhD/NN
CEO/NN
Chapter/NN 4 of the US Commercial Code
more than 120,000 State/NN owned and collective enterprises
person A/NN
grade A/NN eggs

The following are done as NNP:

Islamic/NNP Revolution/NNP
Hakka/NNP eight tone, Fujian/NNP ten tone
Chinese yew (Taxus/NNP mairei/NNP)
Abdullah/NNP bin/NNP al/NNP -/HYPH Aziz/NNP
Chiang/NNP Kai/NNP -/HYPH shek/NNP

2.2.3   Cardinal Numbers in Proper Names: NNP or CD

As with adjectives, cardinal numbers occurring in proper names are tagged NNP.

Formula/NNP One/NNP racing
Shahab/NNP 5/NNP rocket
King Mohammed IV/NNP
F/NNP -/HYPH 16/NNP fighter

2.2.4   Singular vs. Plural Nouns: NN(P) or NN(P)S

Verb agreement is the standard test for distinguishing between singular and plural nouns, despite presence or absence of plural morphology. Truly ambiguous items, such as *data*, are tagged as singular if context does not supply an answer.

the United/NNP States/NNP is
the Phillipines/NNP is
my politics/NNS are my business
linguistics/NN is a dreary subject
the data/NN supports
the data/NNS show
Linguistic Data/NNP Consortium

The following are items that should be tagged NN(P).

> a 2.45 m/NN jump
> economics/NN
> United Nations/NNP
> United Arab Emirates/NNP
> Tianjin Customs/NNP

The following should be tagged NN(P)S.

> sports/NNS channels
> kudos NNS
> personnel/NNS
> police/NNS said that
> the outskirts/NNS of the town
> headquarters/NNS
> cattle/NNS ranch
> jumped 2.45 m/NNS
> foreign affairs/NNS
> savings/NNS and loans/NNS
> communications/NNS
> network sales/NNS income
> cattle/NNS ranch
> Foreign Affairs/NNPS Bureau
> since the Middle/NNP Ages/NNPS
> the Olympic Games/NNPS
> Palestinian/NNP Territories/NNPS
> Alshabura Martyrs/NNPS Camp

2.2.5   Nouns and Adverbs: NN or RB

The most problematic items for this distinction are *home* and *away*.

> at home/NN
> went home/RB
> home/RB game
> away/RB game

2.2.6 Nouns and Present Participles: NN or VBG

Distinguishing between nominal and verbal uses of gerunds is notoriously difficult. No policy changes have been adopted; the following decisions are included to serve as point of reference for current annotators.

The following items are done as NN:

> the shooting/NN
> a firefighting/NN plane
> broadcasting/NN rights
> satellite broadcasting/NN
> the opening/NN ceremony
> operating/NN system
> training/NN times
> fierce fighting/NN
> killed during fighting/NN with the enemy
> a symbolic handing/NN over/RP ceremony
> car racing/NN
> the hijacking/NN operation
> the negotiating/NN table
> international campaigning/NN
> fell to X on closing/NN
> return to playing/NN
> shaving/NN - brush

The following items are done as VBG:

> disturbing/VBG statistics
> competing/VBG teams
> rampaging/VBG camel
> a cheering/VBG crowd
> in spite of increasing/VBG acts of violence
> the qualifying/VBG matches
> working/VBG out
> bird/NN -/HYPH watching/VBG

2.3 Verbs

2.3.1 Participle or Adjective

The distinction between adjectives and present participles is notoriously difficult for annotators to make. The following items have been discussed and decided on, and so are included here for reference.

VBG or JJ   The following items are done as JJ.

> a losing/JJ score
> fighting/JJ spirit
> the missing/JJ include

The following are done as VBG.

> award/NN -/HYPH winning/VBG
> top/RB - /HYPH grossing/VBG
> ruling/VBG party
> best/RBS -/HYPH selling/VBG
> the finishing/VBG touches
> disputing/VBG parties
> opening/VBG arguments
> the stimulating/VBG performance

VBN or JJ   The following are tagged JJ.

> refined/JJ rice
> the mixed/JJ economy
> mixed/JJ ownership economy
> employed/JJ people
> extroverted/JJ agriculture
> specialized/JJ accessory factories
> developed/JJ countries
> advanced/JJ technology
> are concerned/JJ about
> handheld/JJ device
> known/JJ information
> involved/JJ parties

sunken/JJ submarine
wasted/JJ attacks
the injured/JJ
middle - aged/JJ people
feeling depressed/JJ
world - renowned/JJ
human - faced/JJ jar
square - patterned/JJ pottery jar
finished/JJ reproductions
round - faced/JJ
three - legged/JJ
medium - sized/JJ

The following are tagged VBN.

a blown/VBN glass factory
bonded/VBN area
the proposed/VBN project
processed/VBN products
government run/VBN hotel
preferred/VBN shares
booby - trapped/VBN
revered/VBN religious establishments
heavily - policed/VBN area
liquified/VBN natural gas
televised/VBN debate
failed/VBN attempt to assassinate
to stop Libyan/ - backed/VBN
martyred/VBN Palestinian people
repeated/VBN attacks
fast - paced/VBN
man - made/VBN
above - mentioned/VBN

2.3.2   VBG or IN

Gerunds functioning as prepositions are labeled VBG if there is a corresponding verb.
Otherwise, they are tagged IN.

including/VBG

according/VBG to
during/IN

2.3.3 VB or VBP

In potentially ambiguous contexts, verbs are assumed not to have any morphology. '*Thank you*'is analyzed as a pro-drop sentence rather than an imperative, and so *thank* should be tagged VBP.

how come/VB
let/VB alone do that
to let go/VB of
Thank/VBP you very much

2.4 Other Word Classes

2.4.1 Adverbs, Particles and Prepositions

No policy changes have be implemented in distinguishing adverb, particles and prepositions; the items below represent some potentially difficult cases which have cropped up in ETTB.

IN or JJ   The following decisions have been made.

near/IN the house
coming near/JJ their goals
none/NN other/JJ than/IN

IN or RB   The following are tagged IN.

over/IN 10
along/IN with
it is up/IN to/IN you
as/IN though/IN

The following are tagged RB.

about/RB 10
around/RB 10
aside/RB from
away/RB from
was about/RB to
came in/RB second/RB
move around/RB inside/IN Namibia
get close/RB to
live up/RB to its reputation

RB or JJ   The following are tagged JJ.

stand firm/JJ
broadcast the matches live/JJ
released of late/JJ
late/JJ January
do right/JJ by me

The following are tagged RB.

hold a meeting late/RB this month
cut short/RB a visit
the surveillance cameras were on/RB
get right/RB
die young/RB
came in second/RB
came first/RB
at eight sharp/RB

RB or RP   The following are tagged RB.

carry on/RB
the storm died down/RB
to go on/RB until then
calm down/RB
live up/RB to its reputation
play about/RB
let their hair down/RB
spring up/RB

went on/RB to say

The following are tagged RP:

brought about/RP
give in/RP to pressure
set up/RP
sign up/RP for
come up/RP
start out/RP
the plane takes off/RP
take clothes off/RP
play down/RP
a sea mine left over/RP from WWII
oxygen will run out/RP completely
flare up/RP
violence broke out/RP
find out/RP
watch out/RP for/IN X
working out/RP (in both the senses of 'exercising' and 'succeeding')
mellow out/RP
turned out/RP to be
pull out/RP

2.4.2   Non-adverbs Functioning Adverbially

Tokens which are not normally tagged as adverbs are tagged RB* when they function as part of a multi-token ADVP in . In other words, all children of an ADVP should be tagged as some flavor of adverb.

no/RB longer/RB
at/RB all/RB
pretty/RB much/RB went on all/RB day/RB

2.4.3   Symbols: SYM

The SYM tag is used for all non-punctuation non-alphanumeric characters, regardless of function. This includes currency names which include both alphabetical characters and

and symbols. It is also applied to the occasional stray Chinese (or, in principle, any non-Roman) character.

50 %/SYM
NT$/SYM 500

2.4.4 Foreign Words: FW

If, in highly-edited context, the word would be italicized it should be FW.

prima/FW faciae/FW
etc./FW
a/FW priori/FW
je/FW ne/FW sais/FW quoi/FW
terminus/FW ante/FW quem/FW
raison/FW d'etre/FW
charges/FW d'affaires/FW
status/FW quo/FW

Fully integrated prepositions such as *per*, *in* and *vis-a-vis* are tagged as IN, even in collocations with clearly non-English words:

in/IN absentia/FW
per/IN capita/FW
per/IN annum/FW
in/IN absentia/FW
per/IN capita/FW
per/IN annum/FW
vis/IN -/HYPH a/IN -/HYPH vis/IN the agreement

Borrowed nouns are tagged NN, however. Also, *am*, *pm*, *AD*, *AH*, etc. are tagged as nouns.

ma/NN po/NN tofu
saag/NN paneer/NN
a yuhua/NN stone
beiguan/NN and nanhuan/NN music
12 am/NN
2009 AD/NNP

Exclamations from foreign languages are tagged UH rather than FW.

Gesundheit/UH !
Merde/UH !
L'chaim/UH !
Bon/UH appetit/UH !
Ciao/UH !

2.4.5   Wh-words: W*

When a *wh-* word is used in a context in which there is no (associated) *wh-* movement, it is tagged as a wh-word:

I'll do whatever/WP
I'll do whichever/WDT
I can go whenever/WRB
and what/WP not
must comply whatever/WDT the circumstances

2.5   GW and XX in speech data:

The GW tag is used in the Arabic broadcast news corpus. It should be used, as outlined in the guidelines for the Switchboard corpus, to join tokens that have been mistranscribed as separate tokens. The final token has the correct POS tag and all ones to be joined to it are tagged GW.

may/GW be/RB
a/GW round/RB six hundred
not/GW with/GW standing/IN
north/GW west/NN Pakistan

The XX tag is used for incomplete material. If the partial token identically matches the beginning of the following complete token, it is assumed to be restarted instance of that token. Otherwise, the speaker's intended utterance is deemed to be unrecoverable and the token is tagged XX. Expanding on this use, the XX tag is also used for the "partial word" token *%pw* in Arabic broadcast news data.

th-/DT the/DT
un-/XX the/DT
%pw/XX peace/NN

2.6 Punctuation

The following POS tags exist for punctuation:

" Left (double or single) quote
" Right (double or single) quote
-LRB- Left parenthesis (round, square, curly or angle bracket)
-RRB- Right parenthesis (round, square, curly or angle bracket)
, Comma, mid-sentence punctuation (, ; . . .)
. Sentence-final punctuation (. ! ?)
: Colons and dashes (: – -)
HYPH Hyphen in hyphenated items that are split up (London - based)

The period tag should never occur sentence-internally. Comma is the default tag for non-final punctuation. However, sentence-initial * is tagged as SYM.

2.7 Additional Guidelines for POS-tagging Webtext

Two additional POS tags are used for webtext data. These are:

NFP for Non-Final Punctuation
ADD for web ADDresses

NFP is used for superfluous or decorative punctuation, including punctuation occuring at the beginning of a sentence as well as punctuation marking text boundaries or used as 'decoration'. Emoticons are also tagged as NFP, as are punctuation marks that make up an entire sentence unit.

-/NFP Did they pay for them?
People for sale :-)/NFP
_/NFP _____/NFP _/NFP

ADD should be used for URLs and e-mail addresses.

http://www.islamonline.net/Arabic/news/2004-12/05/images/pic05b.jpg/ADD
rayhanenajib@menara.ma/ADD

2.7.1 Punctuation in Webtext

Because of the non-standard use of punctuation in webtext data, tags for punctuation is determined by position in the sentence. Punctuation at the end of a sentence unit (SU)

is tagged as period, regardless of whether it is a period, comma, exclamation. Likewise, sentence-medial punctuation is tagged as comma.

> For the first time after the war ../, cameras are rolling .
> The list goes on ,/.
> Thank you for passing by ../.
> To be continued .../.
> My brother Salam88 :/.

2.7.2 Capitalization

In webtext, capitalization is a less reliable guide as to whether than in newswire and other text genres. Web-specific items, such as web handles, should receive whatever they would in non-web data. For example, a web handle like *@z@b asaah* should be POS-tagged as NNP despite the lack of capitalization.

3 Treebank

3.1 Introduction

This section is a summary of current Treebank policies. It is intended to supplement the most recent published addendum (Bies, Warner, Brisson and Mott 2004). These guidelines reflect policies developed during the annotation of the English-Chinese Treebank and the English-Arabic Treebank (referred to collectively as EXTB), and assume the policies described there (as well as the original manual) as a point of departure. Substantive changes in general annotation policy are described first. Then, as much as possible, changes or elaborations in policy are organized by either the relevant node label or dash tag.

3.2 Changes from Penn Treebank IIa

ETTB relies heavily on the modifications to original Penn Treebank as described by (Ann Taylor ref.). The most important changes are summarized below. More detailed information by Ann Taylor, in particular regarding distinguishing between raising, ECM, subject- and object control constructions, can be found at

`www-users.york.ac.uk/~lang22/TB2a_Guidelines.htm`

.

3.2.1 Copular Verbs

The range of copular verbs is smaller than in previous Treebanks. They have the following argument structure:
single NP-PRD, ADJP-PRD, or PP-PRD complement

Test: can take a bare adjectival complement, but not an infinitival complement. Note that there are four verbs - *seem, appear, prove, turn out* - that take bare adjectival complements but can also take infinitival complements. These were previously annotated as ADJP-PRD, but are now treated as raising predicates.

    remain (uncertain/good friends/IN good shape) [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]
    turn (cold/traitor) [adj/n]
    wind up (drunk/a millionaire) [adj/n]

Some of these verbs can also take PP complements. (Ex: 'feel under the weather') These PPs receive the -PRD function tag.

Examples:

```
(S (NP-SBJ He)
   (VP feels
       (ADJP-PRD annoyed)))

(S (NP-SBJ He)
   (VP feels
       (NP-PRD a fool)))

(S (NP-SBJ He)
   (VP feels
       (PP-PRD under (NP the weather))))
```

3.2.2 Ditransitive verbs

Ditransitive verbs are a larger class than in earlier Treebanks. The have to following argument structure:

NP + S (infinitival) complements; the null subject of the infinitive is always indexed. with the NP object.

```
(VP ask
  (NP-1 him)
  (S (NP-SBJ-1 *PRO*)
     (VP to (VP XYZ))))
```

Comprehensive (but not exhaustive) list:

advise, ask, admonish, allow, authorize, beg, beseech, bribe, catch, caution, challenge, choose, command, commision, compel, convince, counsel, designate, detail, devise, direct, dispatch, empower, enable, encourage, enjoin, entice, exhort, expand, forbid, force, help, hire, impell, implore, incite, induce, influence, inspire, instruct, invite, lead, leave, license, move, name, oblige, order, permit, persuade, pray, press, prod, program, promise, prompt, push, recommend, remind, request, require, retain, schedule, select, sensitize, slate, spur, teach, tell, trust, urge, warn

Test: The NP immediately following the verb cannot be expletive 'there', weather 'it', or an idiom chunk. If it can be, the verb takes a small clause.

Mary expected there to be a party (small clause)
*Mary persuaded there to be a party (ditransitive)
Mary believed it to be raining
*Mary decided it to be raining
Mary found the cat to be out of the bag
*Mary encouraged the cat to be out of the bag

Examples:

Active:

```
(S (NP-SBJ I)
   (VP begged
       (NP-1 him)
       (S (NP-SBJ-1 *PRO*)
          (VP to
              (VP do
                  (NP that))))))
```

Passive:

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

```
      (VP was
          (VP asked
             (NP-1 *)
             (S (NP-SBJ-1 *PRO*)
                (VP to
                    (VP do
                        (NP that)))))))))
```

Reduced Relative:

```
    (NP (NP the person)
        (VP authorized
            (NP-1 *)
            (S (NP-SBJ-1 *PRO*)
               (VP to
                   (VP do
                       (NP that))))))
```

Note that some verbs have more than one argument structure and fall under multiple classes. For example, in the absence of a direct object, 'ask' functions as a monotransitive verb:

```
(S (NP-SBJ-1 He)
   (VP asked
       (S (NP-SBJ-1 *PRO*)
          (VP to
              (VP do
                  (NP that))))))
```

But note the differences between this tree and the passive ditransitive example above.

3.2.3 Verbs with Secondary Predicates

This is a new class. Argument Structure:

```
(VP name
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *)
           (NP-PRD Bob)))
```

*NP + S-CLR complements
S-CLR is non-infinitival, contains a bare NP-PRD, ADJP-PRD, PP-PRD, or ADVP-PRD
null subject of S-CLR coindexed with NP object

25

Label verbs: Non-exhaustive list:

appoint, call, code-name, designate, dub, elect, entitle, headline, label, mark, name,
nickname, quote, rank, rate, rename, subtitle, tag, term, title, vote

Tests:

Label verbs cannot take an infinitival complement (if the verb can, it takes a small clause
or is a ditransitive)

I nicknamed him Bob
*I nicknamed him to be Bob
'nickname' takes an S-CLR secondary predicate
I considered him a fool
I considered him to be a fool
'consider' takes a single S complement

Some label verbs can also be paraphrased with an 'as' PP-CLR:

I appointed Smith press secretary
I appointed Smith as press secretary

Examples:
Active:

```
(S (NP-SBJ I)
   (VP appointed
       (NP-1 him)
       (S-CLR (NP-SBJ-1 *PRO*)
              (NP-PRD president))))
```

Passive:

```
(S (NP-SBJ-1 He)
   (VP was
       (VP appointed
           (NP-1 *)
           (S-CLR (NP-SBJ-1 *PRO*)
                  (NP-PRD president)))))
```

Reduced Relative:

```
(NP (NP the man)
    (VP named
        (NP-1 *)
        (S-CLR (NP-SBJ-1 *PRO*)
               (NP-PRD Bob))))
```

Other Verbs: Non-exhaustive list:

   drive, keep, leave, render, turn

These verbs have a mandatory secondary predicate that describes a resultative state. We
mark this secondary predicative as S-CLR.

Active:

```
(S (NP-SBJ This)
   (VP drives
       (NP-1 me)
       (S-CLR (NP-SBJ-1 *PRO*)
              (ADJP-PRD crazy))))
```

with NP:

```
(VP keep
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *PRO*)
           (NP-PRD an innocent)))
```

with ADJP:

```
(VP keep
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *PRO*)
           (ADJP-PRD sane)))
```

with PP:

```
(VP keep
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *PRO*)
           (PP-PRD in (NP the loop))))
```

with ADVP:

```
(VP keep
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *PRO*)
           (ADVP-PRD away)))
```

with VP:

```
(VP keep
    (NP-1 him)
    (S-CLR (NP-SBJ-1 *PRO*)
           (VP feeling (ADJP-PRD good))))
```

Passive:

```
(S (NP-SBJ-1 I)
   (VP was
       (VP driven
           (NP-1 *)
           (S-CLR (NP-SBJ-1 *PRO*)
                  (ADJP-PRD crazy)))))
```

3.2.4   Other Uses of S-CLR

Distinguishing S-CLR from S-ADV   All secondary predicates coindexed to the object are labelled S-CLR. Secondary predicates coindexed to the subject are labelled S-ADV when the predicate can be paraphrased with *while* or *being*, or be preposed. Most subject-controlled secondary predicates are of the S-ADV type.

> he walked into the room naked
> he walked into the room while/being naked
> naked, he walked into the room

```
(S (NP-SBJ-1 he)
   (VP walked
       (PP-DIR into the room)
       (S-ADV (NP-SBJ-1 *PRO*)
        (ADJP-PRD naked))))
```

"fall silent"   he fell silent
*he fell while/being silent
*silent, he fell

28

```
(S (NP-SBJ-1 he)
   (VP fell
       (S-CLR (NP-SBJ-1 *PRO*)
              (ADJP-PRD silent))))
```

Resultatives

```
(S (NP-SBJ They)
   (VP painted
       (NP-1 the apartment)
       (S-CLR (NP-SBJ *PRO*-1)
              (ADJP-PRD red))))
```

See the "Resultatives" section in the paper manual (PP. 258-260) for more on this pattern.

Infinitives off of "use" verbs: Non-exhaustive list:

use, design, hire, build, write

The infinitives of these verbs are treated as 'semi-complements' and tagged S-CLR. In theory the null subject should always be object-controlled, but in practice can be coindexed with either the subject of the object of the main sentence as appropriate.

Note that this is the only pattern in which S-CLR can be infinitival.

```
(S (NP-SBJ-1 Loews)
   (VP used
       (NP a mechanical process)
       (S-CLR (NP-SBJ-1*PRO*)
              (VP to
                  (VP make
                      (NP filters))))))
```

"Spend time doing X"

```
(S (NP-SBJ-1 He)
   (VP spent
       (NP time)
       (S-CLR (NP-SBJ-1*PRO*)
              (VP doing
                  (NP laundry)))))
```

Note that the manual also presents a treatment of this pattern in which "doing laundry" is an S-NOM adjoined to 'time'. We have discontinued this usage.

3.2.5 Verbs with Single S Complement

This class includes monotransitive, small clause, ECM passive, and raising and subject control constructions.

Argument Structure: Single S Complement (no restrictions on form)

*A large number of verbs that were formerly analyzed as taking small clauses are now treated as either ditransitive or label verbs.

*The following were formerly treated as copular verbs, but now take S complements:

   seem, appear, prove, turn out

Examples:

```
(S (NP-SBJ-1 He)
   (VP seems
       (S (NP-SBJ-1 *)
          (NP-PRD a fool))))

(S (NP-SBJ-1 He)
   (VP seems
       (S (NP-SBJ-1 *)
          (VP to
              (VP be
                  (NP-PRD a fool))))))

(S (NP-SBJ-1 He)
   (VP was
       (VP believed
           (S (NP-SBJ-1 *)
              (VP to
                  (VP be
                      (NP-PRD Bob)))))))


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

```
(VP aim
    (S (NP-SBJ-1*PRO*)
       (VP to (VP please)))))
```

Below are Ann Taylor's non-exhaustive lists. All these verbs can take single S complements. (But not necessarily: note, for instance, that some subject control verbs have ditransitive flavors as well. "I asked to do that" versus "I asked him to do that".)

ECM: *assume, believe, cause, consider, declare, deem, estimate, expect, find, get, hold, imagine, intend, judge, know, make, mean, need, perceive, project, repute, rumor, report, say, see, show, suppose, thin, want, wish*

Small Clause: *believe, consider, declare, deem, fancy, fear, find, get, have, hold, judge, make, presume, proclaim, pronounce, prove, regard, render, report, rule, rumor, see, think, want, wish*

Predicate Raising: *appear, begin, continue, have, fail, happen, need, prove, say, seem, start, stop, tend, apt, certain, due, likely, sure, unlikely, be about to, be bound to, be going to, be supposed to, turn out to, used to*

Subject Control: *act, admit, afford, agree, aim, apply, arrange, ask, attempt, avoid, be willing, bother, call, come, care, choose, claim, clamor, concede, consider, conspire, decide, decline, delay, deny, deserve, determine, disclaim, discuss, end, enjoy, elect, favor, feel, figure, flock, force, forget, five, get, go, hate, help, hesitate, hope, intend, jump, keep, know, lay, learn, like, look, love, manage, mean, mind, miss, motivate, move, negotiate, offer, opt, plan, pledge, plot, pose, ponder, prefer, prepare, press, proceed, profess, promise, propose, push, quite, race, recell, refuse, remain, report, resolve, risk, rule, rush, scramble, seek, serve, set, sign, sound, spend, stand, strive, struggle, suffice, swear, threaten, try, turn, undertake, vote, vow, wait, want, wind, wish, work*

3.3 Parentheticals

The usage of the PRN node is now more restricted than in earlier Treebanks. Formerly, any constituent or group of constituents bounded by brackets was enclosed in a PRN node. Similarly, other punctuation (mainly commas and dashes) could set off a PRN node if an annotator decided that a span was an aside or otherwise not part of the rest of the sentence. Now, PRN is to be used for "sentential interpolations". This means that every PRN should dominate some flavor of an S- or SBAR-node (with certain exceptions outlined below).

```
(NP Bob Smith (PRN -LRB- (ADJP aged 50) -RRB- ))
```

should now be:

```
(NP (NP Bob Smith) -LRB- (ADJP aged 50) -RRB- )
```

with no PRN node and the ADJP adjoined. Note that the brackets should be paired and attached high. This is a departure from the previous policy in which brackets were attached under the PRN node.

There are a few common constructions where PRNs were often used. These are exemplified below:

Renamings-

```
(NP
  (NP (NP (NP the Ministry)
                  (PP of (NP the Interior)))
          -LRB-
          (NP MOI)
          -RRB-)
    and
    (NP
      (NP (NP the Council)
        (PP of (NP Agriculture)))
      -LRB-
      (NP COA)
      -RRB-))
```

Additional NP modifiers:

```
(VP have
  (VP been
    (VP brought (NP-1 *) (ADVP together)
      (PP as
          (NP
            (NP the  (NML Water Resources)  Department)
            -RRB-
            (PP for
              (NP
                (NP
                  (NP the implementation)
                  (PP of
                    (NP water conservancy)))))
            and
            (NP (NP provision)
             (PP of
                (NP water supplies)))))
```

```
                    -LRB-))
```

Additional sentential/VP modifiers:

```
    (S He arrives tomorrow -LRB- (PP-LOC at JFK) -RRB- .)
```

PRN is no longer an allowable top node (see 3.5 above). This applies to material in parentheses in headlines of Sinorama articles:

```
    (FRAG -LRB-
      (NP Eric Lin)
      /
      (NP (NP photos)
        (PP by
          (NP Diago Chiu)))
      /
      (VP tr. (NP *)
        (PP by
          (NP-LGS Scott Williams)))
      -RRB-)
```

The instances where the usage of PRN is still sanctioned can be divided into two categories. 1. Sentential interpolations: This is most common with discontinuous quotations, such as the following:

```
    (S '' I love '' (PRN , says Bill ,) '' peaches .'')
```

Other interpolations occur as well:

```
    (S It should rain tomorrow (PRN -LRB- God willing -RRB-) .)
```

This general policy is suspended if there is no alternative to incorporate the constituents into the rest of the sentence in any sane way. In such cases, annotators should leave them under PRN node. In this use, the PRN node is regarded as "extra-syntactic".

```
    (S
      (S-NOM-SBJ (NP-SBJ*PRO*)
        (VP having (NP one)))
      (VP was
        (ADJP-PRD so expensive)
        (PRN -
          (FRAG
            (NP
              (NP (QP NT$   70 - 80,000) *U*)
              (PP for  (NP a phone)))
```

```
           ,
           and
           (ADVP
             (ADVP as much)
             (PP as
               (NP
                 (NP (QP NT$ 100,000) *U*)
                 (PP in (NP total)))))
           (SBAR once
             (S (NP-SBJ you)
               (VP added (PRT in)
                 (NP
                   (NP (NML set - up) fees)
                   ,
                   (NP a deposit)
                   and
                   (NP connection costs))))))

       -)))

   (NP little (PRN , (SBAR if (FRAG (NP any)))) resolve)
```

Other non-sentential uses of PRN include:

1. Locations in company names:

```
       (NP Toshiba  (PRN -LRB- (NP Dalian) -RRB- ) Company)
```

2. Temporal noun phrases signifying duration of a period:

```
       (NP the Qing dynasty
         (PRN -LRB- (NP-TMP (NP 1644)
                     (PP -
                       (NP 1911)) -RRB-)))
```

3. Phrases relating items in the text to items outside the text, such as photographs:

```
       (NP Bob Smith)
       (PRN (ADVP above))
```

3.4   Speech

The guidelines outlined here for speech effects are based on those developed for the Switch-board corpus (REF), and the policies and decisions discussed here are intended to clarify

and expand on those. Annotated speech data include two important features that are not used for other types of data: the EDITED node and the -UNF tag.

If a restart has the same initial letter(s) as the following token, give it the same node/POS information.

Example: "w- well you know you just"

```
(S-UNF (EDITED (INTJ (w-))
       (INTJ well)
       (PRN (S (NP-SBJ you)
               (VP know)))
       (NP-SBJ you)
       (ADVP just))
```

On the same note, do not read information that is not there into unfinished tokens. This means liberal use of X and the POS tag XX (these are always paired) when you're not sure what the token is.

"w- di- he didn't"

```
(S-UNF (EDITED
(S (EDITED (X w-)) (EDITED (SQ di- )) (NP-SBJ He) (VP did 'nt))
```

An EDITED node never dominates an entire utterance. A -UNF tag is placed on the appropriate node instead. (One corollary of this is that is can thus never be a top node, see also 3.5).

```
(S (INTJ so)
   (NP-SBJ she)
   (VP 's
       (NP-PRD-UNF a g-)))
```

Traces are placed inside of EDITED nodes when there is enough structure in the disfluency to allow them to appear in their proper position. Otherwise, they are not indicated, as in the following.

```
(SBARQ (EDITED (SBARQ (WHNP what)
                      (SQ-UNF 's)))
       (WHNP-1 what)
       (SQ 's
           (NP-SBJ toning)
           (NP-PRD-1 *T*)))
```

3.4.1 UNF

The "unfinished-ness" that is marked by the UNF tag applies not only to the constituent it is marked on, but also to all parent nodes of that constituent. So, in the following, the the VP and S are both unfinished by dint of the UNF tag on the direct object.

```
(S (NP-SBJ I)
    (VP saw
        (NP-UNF the ) ))
```

The second incomplete token does not get an EDITED because of the final position in the utterance. We assume the whole utterance is incomplete instead.

```
(S (EDITED (NP-UNF that-) (INTJ-UNF ye-))
```

%pws (i.e., the tokens representing the translation of partial words in a language other than English) are treated as punctuation. That is, they are left high in the the tree. In general, this means they are often outside of EDITED nodes

%pws can, however, go under EDITED nodes if and only if the annotator has a very strong intuition that the incomplete material the %pw represents would unambiguously be in the restart if it were completed.

```
(EDITED (NP-UNF the %pw )) (NP the elephant )
```

3.4.2 Welcome(...)

Welcome should treated as INTJ when a) it occurs by itself or b) occurs with only a vocative NP:

```
(INTJ Welcome)
(INTJ (INTJ Welcome) (NP-VOC Dorothy))
```

Otherwise, it should be treated as VP; if there is no overt subject, it is treated as pro-drop:

```
(S (NP-SBJ *PRO*)
    (VP Welcome
        (NP Dorothy)
        (PP-CLR to Munchkin Land)

(S (NP-SBJ *PRO*)
    (VP Welcome
```

```
            (ADVP back)
            (PP-CLR to the program ))).)
```

3.4.3   Thank you

Thank you is treated as pro-drop, rather than as an imperative

```
    (S (NP-SBJ *)
        (VP Thank
            (NP you) ).)
```

3.4.4   Speaker IDs

In cases where translators have included speaker IDs in the English text, they should be labeled NP. They should not be included in the sentence.

"[speaker4] While participating delegations are about to arrive, planes carrying international aid have continued to land at Jakarta Airport."

```
    (NP [speaker4] )
    (S While participating delegations... )
```

3.5   Valid Top Nodes

The only two node labels that are explicitly disallowed from being top nodes are NML, VP and PRN. For PRN, see **??** above. The invalidity of NML as a top nodes falls out of its regular distribution (it only occurs inside NPs).

A FRAG is inserted over the VP node to indicate that there is missing material, including by definition a syntactic subject. This decision was largely made to facilitate Propbank annotation, in which all verbs should be linkable to their arguments.

```
    (FRAG (VP xxx)
    *(VP xxx)
```

3.6   Node Labels

3.6.1   NPs

Dates and Place Names   Dates and place names: In place names such as "San Francisco, California" and dates such as with the final NP(s) getting -TMP or -LOC as appro-

priate.

```
(NP (NP San Francisco)  ,
    (NP-LOC California))

(NP (NP Portland)    ,
    (NP-LOC Maine))

(NP (NP next Monday)   ,
    (NP-TMP October 23))

(NP (NP Monday)  ,
    (NP-TMP October 23rd))

(NP (NP Monday)
    (NP-TMP the 23rd))

(NP (NP October 23)  ,
    (NP-TMP 2006))

(NP (NP (NP Monday)      ,
        (NP-TMP October 23))   ,
    (NP-TMP 2006))
```

Patterns like "9am local time" are done as:

```
(NP-TMP (NP 9 am)
        (NP-ADV local time))
```

ADJP or NP   PPs should not take ADJP complements. As a result, the following are done with NP:

```
(PP in (NP particular/general/common))
(PP at (NP present))
(PP as (NP such))
```

One outgrowth of this as that to avoid PP + ADJP combinations, things like *as soon as possible* are done with an SBAR and a FRAG.

```
(ADVP (ADVP as soon) (SBAR as (FRAG (ADJP possible))))
```

For label verbs cases with 'as', however, the use of ADJP is sanctioned:

```
(VP described (NP it)  (PP-CLR as (ADJP easy))
```

38

Modification of NP-VOC NP-VOCs can be modified, as in the following:

```
(S
  (NP-VOC (NP Mr. Ziyad al Ajouz)

          ,
          (PP-LOC in (NP Lebanon)))

  ,
  (NP-SBJ *PRO*)
  (VP thank (NP you)
    (ADVP very much))
  .)
```

Recursive Adjunction of Relative Clauses Relative clauses, both reduced and full, restrictive and non-restrictive, are now recursively adjoined to an NP that includes everything that comes before them:

```
(NP (NP (NP the President)
        (PP of
            (NP the United States)))
    (SBAR (WHNP-1 who)
          (S .... )))

(NP (NP (NP the book)
        (PP about
            (NP toads)))
    (SBAR that is on the table)

(NP (NP (NP the book)
        (VP read
            (NP *)
            (PP-LGS by (NP me))))
    (SBAR that is on the table))
```

Note that PPs following a relative clause are also recursively adjoined:

```
(NP (NP (NP the book)
        (SBAR that is on the table))
    (PP about toads))
```

PP Premodifiers of NP PPs which occur as premodifiers inside NPs. The most common of these are *among them* and *for example*

```
(NP (NP 11 people) (NP (PP among (NP them)) 7 soldiers))
(NP (NP Arab countries) (NP (PP for (NP example) Syria)))
```

NAC adjoined to NP   Since the removal of the coordinating conjunction does not affect the meaning, this is a valid use of NAC.

```
(NP (NP his loss) (NAC and (PP in ...)))
```

3.6.2   ADVP

Focus Particles   Despite being pos-tagged RB, *just, even,* and *only* are treated as focus particles.

In NP/QP/ADJP/ADVP they do not project an ADVP:

```
( (S (NP-SBJ (RB only) (DT the) (JJ assistant) (NN manager))
    (VP (MD can)
        (VP (VB talk)
            (PP-CLR (IN to)
                    (NP (DT the) (NN manager))))))
 )

( (S (NP-SBJ (DT the) (JJ new) (NN park))
    (VP (MD will)
        (VP (VB have)
            (NP (QP (RB only) (CD 1,500))
                (NNS spaces)))))
 )

( (S (NP-SBJ (RB even) (JJ big) (NNS players))
    (VP (VBP are)
        (RB n't)
        (ADJP-PRD (JJ immune)
                (PP (IN to)
                    (NP (NP (DT the) (NNS rigors))
                        (PP (IN of)
                            (NP (NN program) (NN trading))))))))
 )

( (S (NP-SBJ (NP (PRP it))
            (SBAR (-NONE- *EXP*-1)))
    (VP (VBZ 's)
        (ADJP-PRD (RB even) (JJ possible))
        (SBAR-1 (-NONE- 0)
                (S (NP-SBJ (NNP Mr.) (NNP Morita))
                    (VP (MD would)
                        (VP (VB have)
                            (VP (-NONE- *?*)
```

```
                                 (ADVP (RB too)))))))))))
```

In SBAR and PP, *just, even, only* project an ADVP, but are still inside the following
phrase.

```
    (ADJP (JJ fit)
        (PP (ADVP (RB only))
            (IN for)
            (NP (NNS foreigners))))


    (SBAR-ADV (ADVP (RB even))
            (IN though)
            (S (NP-SBJ (NNP NATO) (NNS negotiators))
                (VP (VBP have)
                    (NP (NP (QP (RB only) (CD 10))
                            (NNS months))
                        (VP (VBN left)
                            (NP (-NONE- *))))
                    (PP-TMP (IN under)
                            (NP (DT the) (NNP Bush) (NN timetable))))))


    ( (S (NP-SBJ (DT An) (NN oncogene))
        (VP (MD may)
            (VP (VB turn)
                (NP-3 (VBG proliferating) (NNS cells))
                (S-CLR (NP-SBJ (-NONE- *PRO*-3))
                        (ADJP-PRD (JJ malignant)))
                (SBAR-TMP (ADVP (RB only))
                        (RB after)
                        (S (NP-SBJ-1 (DT the) (JJ tumor-suppressor) (NN gene))
                            (VP (VBZ has)
                                (VP (VBN been)
                                    (VP (VBN damaged)
                                        (NP (-NONE- *-1)))))))))
    )
```

But in S and VP *just, even, only* act like other adverbs and attach high.

```
    ( (S (NP-SBJ (PRP they))
        (ADVP (RB just))
        (VP (VBP drift)
            (ADVP-DIR (RB away))))
    )
```

Sentential ADVPs   ADVPs attach at the highest possible level, S or VP (or at conjunction level); essentially ADVPs before the verb attach to S (or a higher VP) and those following attach to the VP containing the verb.

```
(S (NP-SBJ this book)
   (VP is
       (ADVP generally)
       (VP liked
           (PP by
               (NP-LGS children)))))
```

But in VP reduced relatives, ADVPs are inside the VP.

```
(NP (NP a book)
    (VP (ADVP generally)
        liked
        (PP by
            (NP-LGS children))))
```

Non-sentential ADVPs   ADVPs that modify PP or SBAR appear as sister of the P and always project an ADVP.

```
(S (NP-SBJ (DT These) (CD three) (NNS countries))
   (VP (VBP are)
       (RB n't)
       (PP-PRD (ADVP (RB completely))
               (IN off)
               (NP (DT the) (NN hook)))
       (, ,)
       (ADVP (RB though)))
   (. .))

( (S (NP-SBJ-69 (DT The) (JJ Czech) (NN dam))
     (VP (MD ca)
         (RB n't)
         (VP (VB be)
             (VP (VBN operated)
                 (NP (-NONE- *-69))
                 (PP-TMP (ADVP (RB solely))
                         (IN at)
                         (NP (NN peak) (NNS periods)))
                 (PP-MNR (IN without)
                         (NP (DT the) (NNP Nagymaros) (NN project))))))
     (. .))
```

42

```
(S (NP-SBJ (-NONE- *T*-2))
  (VP (VBP are)
      (ADJP-PRD (JJ objectionable))
      (SBAR-PRP (ADVP (RB merely))
                (IN because)
                (S (NP-SBJ (PRP they))
                   (VP (VBP conflict)
                       (PP-CLR (IN with)
                               (NP (PRP$ his) (NN policy) (NNS objectives)))))))))))

(S (NP-SBJ Taiwan)
  (VP should
    (VP have
      (NP opportunities
        (S (NP-SBJ *PRO*)
           (VP to
             (VP enter
               (NP this field)
               (PP (ADVP not merely) as
                 (NP a contract manufacturer)))))))))))
```

"Floating" all The original guidelines allow for "all" to float at VP level without an ADVP parent node

```
(S (NP-SBJ We)
    (VP were
        all
        (ADVP-LOC-PRD there
```

In some instances, this floating *all* cannot be left at VP level. In these cases, it is not labeled ADVP and is left inside the lower node.

```
(S (NP-SBJ He)
  (VP must
    (VP do
      (NP X Y and Z)
      (SBAR-TMP all before time runs out))))

(NP (NP old folk and youngsters) (VP all expecting red envelopes of cash))
```

Adverbs In NP There are two major cases of adverbs in NPs:

1. the NAMELY/etc case ("elaboration")

2. where the adverb modifies something in the NP (usually a quantifer-type thing)

TYPE 1 The adverb with ADVP label is in the NP. PP/SBAR

```
wsj_1574
(NP (NP (NN interest) (NNS rates))
    (SBAR (WHNP-2 (WDT that))
          (S (NP-SBJ (DT the) (NN company))
             (VP (VBZ guarantees)
                 (NP-2 (-NONE- *T*)))))
    (-LRB- -LRB-)
    (NP (ADVP (RB usually))
        (QP (CD 4) (SYM %) (IN to) (CD 4.5) (NN %))
        (-NONE- *U*))
    (-RRB- -RRB-))

(NP-2 (NP (NP (DT the) (NN bulk))
          (PP (IN of)
              (NP (JJ bad) (NNS assets))))
      (, ,)
      (NP (ADVP (RB mostly))
          (JJ real)
          (NN estate))
      (, ,))
```

A lot of these are pre quantifier/measure words and the adverb really modifies the quantifier. I'm not sure if current policy is (a) or (b)

a. (NP (ADVP (RB nearly)) (PDT all) (DT the) (NNS girls))

b. (NP (RB nearly) (PDT all) (DT the) (NNS girls))

No matter... No matter how big the problem the A-Team is there

```
(ADVP no matter (SBAR-1 how big) (S (NP-SBJ the problem) (VP *?* (ADJP-PRD-1 *T*))))
```

3.6.3 ADJP

"aged X years"

```
(NP (NP 10 men)
    (ADJP aged
```

```
        (NP (QP 34 - 85) years)))
```

"get rid of"

```
    (VP get (S (NP-SBJ *) (ADJP-PRD rid (PP of ...))))
```

'the first to eighth floors'

```
    (NP the (ADJP first (PP to (ADJP eighth))) floors)
```

Consecutive adjectives An ADJP is not inserted for serial adjectives inside an NP unless a comma is present, in which case

```
    (NP her big brown eyes)
    (NP the (ADJP big, friendly) dog)
```

3.6.4 QP

"going on X years"

```
    (PP-TMP For  (NP (QP going  on  ten)  years ) (ADVP now))

    ''tens of thousands of ...''
    (NP (NP tens) (PP of (NP (NP thousands) (PP of (NP Palestinians)))))
```

Ratios are treated as complex quantities.

```
    (NP a (QP 1 : 50) dilution
```

3.6.5 S

'make do' This will be treated as an S complement with an empty subject. This 'do' is understood as an intransitive verb meaning "suffice". Also, the *with*-phrase should be off VP.

```
    (VP make
     (S  (NP-SBJ *PRO*)
         (VP do))
     (PP with ...))
```

"bring X to bear"

```
        (VP bring (NP its diplomatic superiority)
                  (S-CLR (NP-SBJ *PRO*) (VP to ( VP bear))))
```

"be that as it may..."

```
        (S (SBAR-ADV
             (SINV Be
               (NP-SBJ that )
               (SBAR-PRD as
                  (S (NP-SBJ it )
                     (VP may (VP *?*)))))))  ,   ...)
```

"easy come, easy go"

```
        (NP This (NML (VP (VP (ADVP easy) - come) ,
                          (VP (ADVP easy) - go )))   attitude )
```

"have yet to..."

```
        (VP have
          (ADVP-TMP yet )
          (S (NP-SBJ *)
             (VP to open any formal mechanism )))
```

S-NOMs can undergo it-extraposition, as in the following:

```
        (S
             (NP-SBJ (NP It) (S-NOM-2 *EXP*))
             (VP drove (NP-1 me)
               (S-CLR (NP-SBJ-1 *PRO*) (ADJP-PRD nuts))
               (S-NOM-2 (NP-SBJ *PRO*)
                  (VP watching
                    (S
                       (NP-SBJ the
                         (NML (NML daughter) -
                           (PP in - (NP law))))
                       (ADVP just)
                         (VP sit (ADVP-LOC there))))))))). )
```

3.6.6   SBAR

make X known
```

```
(SBAR-TMP
  soon
  after
  (S (NP-SBJ four Saudi groups)
     (VP made
         (S (NP-SBJ (NP *)
                    (SBAR *EXP*-1))
            (VP known
                (SBAR-1 0
                        (S they were setting up...)))))))))
```

'listen to ...'

This appears to be a unique pattern. It's perfectly valid English, but hard to parse. The least egregious treatment seems to be to treat 'to' as a complementizer.

```
(S (NP-SBJ The bishops)
   (VP listened
       (SBAR to
             (S (NP-SBJ victims)
                (VP break
                    (NP the silence)))))))
```

Function tags on relative clauses Relative clauses do not receive function tags. Note the bare SBAR (instead of SBAR-LOC or SBAR-TMP) in the following examples:

```
(NP (NP the place)
    (SBAR (WHADVP-1 where)
          (S (NP-SBJ he)
             (VP found
                 (NP it)
                 (ADVP-LOC-1 *T*)))))
```

```
(NP (NP the 1960s)  ,
    (SBAR (WHADVP-1 when)
          (S (NP-SBJ the revolution)
             (VP occurred
                 (ADVP-TMP-1 *T*)))))
```

...as were...

```
(NP-SBJ-1 A German journalist)
(VP was
    (VP freed
      (NP *-1)
      (SBAR-ADV as
```

```
        (SINV were
            (NP-SBJ two Filipinos)
            (VP *?*))))
```

3.6.7 VP

Phrasal verbs In instances where phrasal verbs are split into separate tokens, they are
annotated flat.

```
(S (NP-SBJ They)
   (VP have
     (VP been
       (VP job - hopping ...)))))

 (S (NP-SBJ-9 *T*)
        (VP were
           (VP hand - built
              (NP-9 *))) ...)
```

3.6.8 FRAG

"Easier said than done"

```
(FRAG
 (ADJP
   (ADJP Easier
     (S (NP-SBJ-2 *PRO*)
        (VP said (NP-2 *))))
   (SBAR than
     (S (NP-SBJ-1 *PRO*)
        (VP done (NP-1 *)))))
 ,
 (ADVP unfortunately)
 .)
```

For the use of FRAG in headlines see 3.7.6 above.

3.6.9 RRC

```
(NP (NP a book) (RRC (NP 8 inches) (PP in size)))
```

3.6.10 UCP

When a predicate adjective is coordinated with a VP, the -PRD tag is placed on the UCP parent. (UCP-PRD (ADJP) (VP))

3.7 Function Tags

Adverbial dash tags (i.e., LOC TMP, MNR,) are more restricted in their usage than previously. Their use can be described as being literal; -LOC, for example, is only used when it refers to an actual location in real, physical space which can be pointed at (or otherwise indicated).

3.7.1 DIR

When verbs of motion are used non-literally, they receive CLR instead of DIR.

```
going to school
moving to Chicago

going to school (= attending school)
moving to the next topic
```

3.7.2 CLR

Only PP, NP, ADVP and S nodes can be tagged as CLR. Several instances of SBAR-CLR have cropped up in the past, but these appear to be errors and/or based on older policy.

PP-CLR There are several tests that can be applied to determine if a PP should get a CLR tag:

1. The PP cannot be deleted without changing the meaning of the verb.

2. The verb + PP can be paraphrased as a single verb + direct object

   ask for money~requested money
   participate in a meeting~attended a meeting

3. PPs off of verbs of motion when non-physical movement is implied should take CLR rather than DIR: going to school

4. All 'of' phrases off a VP should get CLR.

```
(VP died (PP-CLR of (NP consumption))
(VP remind (NP me) (PP-CLR of that))
```

For use of CLR with verbs of motion, see 3.7.1 above.

Finally, the following collocations were decided to require PP-CLR, and are included here for reference

meet with
engage in diplomatic relations
admit to the UN
participate in
hint at
recover from
exclude from
combine with
connect to
divide into four countries
differs from
hope for

NP-CLR   Some examples of where NP-CLR is used include the following.

```
(VP (NP take (NP-CLR advantage) (PP-CLR of ...)))
(VP gain (NP-CLR control) (PP-CLR of ...))
```

ADVP-CLR   The most common examples of ADVP-CLR are in rankings in descriptions of sporting events (see 3.9.1). It is also used in financial speak (as outlined in original manual).

"took unusually long" The -CLR tag should be used on the ADVP to maintain the parallelism with "took five years'.

```
(S
 (NP-SBJ the cabinet reshuffle)
 (VP took
   (ADVP-CLR unusually long))
```

put himself up as
do away with
put right/forward/aside/together

looking back
die down
take long
had better
pass away
give in to
pull down
go about one's business
go back (four years)
come across X
look down on X
go public
make good
writ large
come first
give away
go in for
stand out
chip in X
went on to X
cut away X

S-CLR For uses of S-CLR, see the relevant sections on secondary predicates (3.2.3), resultatives (3.2.4), label verbs (3.2.3), "use" verbs (3.2.4

3.7.3 PUT

The -PUT tag has been retired. Instead, all ADVPs and PPs that are second "objects" of *put* receive -CLR; those that are locatival *senso strictu* are tagged -LOC as well.

(VP put (NP the book) (PP-LOC-CLR on the table))

(VP put (NP him) (PP-CLR to the test))

3.7.4 NOM

S vs. S-NOM If verb can take a noun but not a verb complement, then the -NOM tag is required.

We are following the original guidelines regarding the distribution of the -NOM tag with S and SBAR. So, -NOM is never used on the complement of verbs which can take an S(BAR) complement:

If I knew then (SBAR what I know now). The water started (S pumping).

The copula, however, is exempt from this: This is (SBAR-NOM-PRD what I want)

For use of the NOM tag in topicalized structures, see **??** below.

3.7.5   TTL

In addition to the usage outlines in (ref in printed manual), the TTL has slightly extended functions in ETTB. In addition to marking the titles of printed works, it is also used to mark quoted speech that is

In this regard, it marks a constituent and being 'nominal'. It should also be noted that this extension was motivated

```
(

S (CC And )
    (NP-SBJ (PRP I ))
    (VP (VBP quote )
        (NP (NP (DT the )
                (NN maxim))
            (, , )
            (SBARQ-TTL (SBAR-ADV (IN If )
                                 (S (NP-SBJ (DT the )
                                            (NN paterfamilias ))
                                    (VP (VBZ is )
                                        (NP-PRD (DT a )
                                                (NN tambourine )
                                                (NN player)))))
                       (, , )
                       (WHNP-9 (WP what ))
                       (S (NP-SBJ-9 (-NONE- *T*))
                          (VP (MD would )
                              (VP (VB be )
                                  (NP-PRD (NP (DT the )
                                              (NN household)
                                              (POS 's ))
                                          (NN ritual))))))))))
    (. ?)
    (. .)
```

```
(NP-SBJ-9 (NP (DT The )
              (NN slogan))
          (, , )
          (FRAG-TTL (INTJ (UH no ))
                    (PP (IN to )
                        (NP (NN corruption)))))
```

Despite the fact that TTL implies nominal-ness, it can cooccur with the NOM tag.

```
(NP (NP the title )
    (S-NOM-TTL ('' ") (NP-SBJ *PRO*)
                      (VP Getting
                         (PP out
                            (PP of
                               (NP the Economic Crisis)))) "))
```

Titles with Possessives   Titles occurring with possessives are done as follows.

```
(NP (NP Wang Ting-chun  's )
    (NML-TTL  Open Life ))
```

3.7.6   HLN

Any sentence getting the tag should 1) be in 'headline-ese' (missing auxilliaries, determin-
ers, etc.) and 2) be the first sentence about the topic at hand. This section also deals
with issues of somewhat incomplete sentences, and so touches on deciding between S* and
FRAG.

President Clinton to leave

```
(FRAG-HLN (NP-SBJ-1 Clinton) (S-PRD (NP-SBJ *PRO*-1) (VP to leave)))
```

3 soldiers killed by bomb

```
(FRAG-HLN (NP-SBJ-1 3 soldiers) (VP killed (NP *-1) (PP by (NP-LGS bomb))))
```

Syria / Damascus calls for ...

```
(FRAG-HLN (NP-LOC Syria) (S ..... ))
```

```
(PP-HLN to (NP the participants))
```

In broadcast news data, multiple -HLN tags can occur in the same file.

If the headline consists of an easily identifiable (and grammatical) syntactic node, then label it as such, with the -HLN tag:

```
(PP-HLN To (NP The Participants))

(NP-HLN Miscellaneous News From The World's Stadiums)
```

FRAG is used-

1. For 'sentences' in headlinese which are missing auxilaries-

```
(FRAG-HLN (NP-SBJ-1 Clinton) (S-PRD (NP-SBJ *PRO*-1) (VP to leave)))

(FRAG-HLN (NP-SBJ-1 3 soldiers) (VP killed (NP *-1) (PP by (NP-LGS bomb))))

(FRAG-HLN (NP OPEC President)
          (PP-LOC in Tehran)
          (PP-PRP for Talks on the Oil Market))


(FRAG-HLN (NP-SBJ American anti-sanctions group)
          (S-PRD to visit Basra))
```

2. To hold together unlike categories

```
(FRAG-HLN  (NP Austrian Grand Prix)
           :
           (S Irishman Irvine Pulls Out))
```

The -HLN tag is not used on datelines.

```
(FRAG (NP-LOC Moscow) (NP-TMP 7-15) (NP (  AFP) ))
```

Multiple unrelated items are annotated as if they as separate sentences. This is prevalent when there are several articles sharing the same headline.

3.7.7   Ordering of Function Tags

In general, grammatical functions tags should be the rightmost function tag.

LOC-PRD
not: PRD-LOC

3.7.8 Interaction of Function Tags and Movement

When S-NOM and SBAR-NOM undergo movement, their traces should have the same
node label (including the -NOM tag) as they do, rather than being represented as NPs.
Examples of the various types of movement are given below:

Passive traces:

```
(S (SBAR-NOM-SBJ-1 What I said)
   (VP was
     (VP ignored
       (SBAR-NOM-1 *))).)
```

Topicalization:

```
(S (S-NOM-TPC-1 Kicking puppies)
   (NP-SBJ I)
   (ADVP really)
   (VP like
     (S-NOM-1 *T*)))
```

Right Node Raising:

```
(S (NP-SBJ I)
   (VP (VP thought
         (ADVP more)
         (PP about
           (S-NOM-1 *RNR*)))
       and
       (VP regretted
         (S-NOM-1 *RNR*))
       (S-NOM-1 kicking puppies)).)
```

Dash-tags in topicalized structures   If the function tag relates to the internal struc-
ture of the topicalized constituent, it goes on that constituent rather than on the trace.
These tags are -IMP, -CLF, -TTL, and -NOM.

```
(S (S-IMP-TPC-1 (NP-SBJ *PRO*) (VP go (ADVP away)))
   (NP-SBJ He)
   (VP said (S-1 *T*)))
```

```
(S (SBAR-NOM-TPC-3 (WHNP-1 what )
                   (S (NP-SBJ-2 it )
                      (VP does n't
```

```
                                        (VP need
                                            (S (NP-SBJ-2 *)
                                                (VP to
                                                    (VP control (NP-1 *T*)))))))))) ,
                    (NP-SBJ it )
                    (VP should n't
                        (VP control
                            (SBAR-NOM-3 *T*))))
```

If the function tags relates to the topicalized constituent's role in the larger sentence, the tag goes on the trace only. This covers argument (-SBJ, -PRD, -DTV, -BNF, -LGS, -CLR) and adverbial (-ADV, -TMP, -LOC, etc.) function tags.

```
    (SINV (ADJP-TPC-1 Charming)
          (VP are
              (ADJP-PRD-1 *T*))
          (NP-SBJ (NP the works)
                  (PP of (NP Chopin, Schubert, Liszt, and Rafael))))
```

Adverbial tags in these constructions are usually found in conjunction with an argument tag like -PRD:

```
    (S (ADVP-TPC-1 There)
       (NP-SBJ he)
       (VP is (ADVP-LOC-PRD-1 *T*)))
```

3.8   Co-Indexation

3.8.1   Raising/Subject Control in S arguments of verbs

If a verb takes a single S complement that has a null subject, that null subject must be coindexed with the main subject of the larger sentence:

```
    (S (NP-SBJ-1 I)
       (VP want
           (S (NP-SBJ-1 *PRO*)
              (VP to
                  (VP do (NP this))))))
```

The null subject of an S complement of a predicate adjective is similarly coindexed:

```
    (S (NP-SBJ-1 I)
       (VP was
           (ADJP-PRD able
```

```
                                  (S (NP-SBJ-1 *PRO*)
                                     (VP to
                                         (VP do (NP this)))))))))
```

cf: Ditransitive verbs take an NP and S complement. The null subject of this S must be
coindexed, but you can choose to coindex it with the subject or object NP, depending on
which is semantically appropriate.

```
    cf: Existential there is not coindexed:
    (S (NP-SBJ there)
       (VP appears
           (S (NP-SBJ *) (VP to (VP be (NP-PRD a problem))))))
```

3.8.2   Subject Control in S-ADV

Indexation of the the null subject of an S-ADV (and S-PRP, S-MNR, and other adverbial
flavors of S) is not mandatory, but the null subject should be indexed with the subject
of the parent S when semantically appropriate. (Note that the null subject of an S-ADV
should not be coindexed with an NP object of the parent verb, only the subject.)

```
    S (NP-SBJ-1 he)
       (VP walked
           (PP-DIR into the room)
           (S-ADV (NP-SBJ-1 *)
                   (ADJP-PRD naked))))

    (S (NP-SBJ-1 Australia)
       (VP won
           (NP the cup)
           ,
           (S-ADV (NP-SBJ-1 *PRO*)
                   (VP beating (NP Brazil)))))


    (S (S-ADV (NP-SBJ-1 *PRO*)
              (VP feeling (ADJP-PRD tired)))
       (NP-SBJ-1 I)
       (VP entered (NP the room)))


    (S (NP-SBJ-1 We)
       (VP met
           (S-PRP (NP-SBJ-1 *PRO*)
                   (VP to (VP discuss (NP important topics))))))
```

3.8.3 Subject or Object Control in S-CLR Resultatives/Secondary Predicates

The null subject of an S-CLR marking a resultative or secondary predicate can be coindexed with either the subject or NP object of the parent S:

```
(S (NP-SBJ-1 he)
   (VP fell
       (S-CLR (NP-SBJ-1 *PRO*)
              (ADJP-PRD silent))))

(S (NP-SBJ I)
   (VP painted
       (NP-1 the room)
       (S-CLR (NP-SBJ-1 *PRO*)
              (ADJP-PRD red))))
```

3.8.4 Indexation across PP/SBAR

```
(S (NP-SBJ I)
   (VP kept
       (NP him)
       (PP from (S-NOM (NP-SBJ *PRO*)
                       (VP doing (NP that))))))

(S (NP-SBJ I)
   (VP called
       (PP-CLR on (NP him))
       (S (NP-SBJ *PRO*) (VP to (VP do (NP that))))))


(S (NP-SBJ This)
   (VP explodes
       (SBAR-TMP (WHADVP-1 when)
                 (S (NP-SBJ-1 *PRO*)
                    (VP combined
                        (NP-1 *)
                        (PP-CLR with (NP that))
                        (ADVP-TMP *T*))))))
```

3.8.5 Indexation in Reduced Relatives

```
(NP (NP the book)
    (VP given
```

```
(NP *)
(NP me)
(PP by (NP-LGS him))))
```

3.9   Miscellany

3.9.1   Sports and Rankings

"come in first, second third, etc." takes the -CLR tag

```
(VP came (ADVP in) (ADVP-CLR second))
(VP seeded (ADVP-CLR fourth))
(VP ranked (ADVP-CLR fourth))
```

it lost the game to remain without points (S-ADV)

```
(VP lost
        (NP the game)
        (S-ADV (NP-SBJ *PRO*)
                     (VP to (VP remain (PP-PRD without (NP points))))))
```

Scores of the format "1-2" are generally marked as NP-ADV. it lost 25-28 to Qatar

```
(VP lost
        (NP-ADV (NP 25)
                (PP -  (NP 28)))))
```

The exception is when they are prenominal position:

```
(NP the (NML (NML 1) (PP to (NP 1 ))) tie)
```

When 'to' (or the dash representing it) are missing, they are done as NP-ADVs:

"winning three nil"

```
(VP winning (NP-ADV (NP three ) (NP-ADV nil) ))
```

3.9.2   Attachment of Punctuation

The default rule for all punctuation inside a top node is to attach it high, i.e., as far up in the tree as is possible. For more information on parentheses see **??** above. In the Xinhua section of ECTB, paired punctuation was generally attached at the same level; this was

59

discontinued in later corpora except in the case of paired brackets, which still attached at the same level.

3.9.3 Annotator Comments

Annotators are allowed to comment on problematic areas that they encounter.

TRANS: Translation issues occasionally crop up when dealing with source materials originally written in other languages (notable for us, in Chinese and Arabic). The guidelines for annotators are:

1. Do not second-guess unless you have to. If it is possible to interpret the sentence with the words that are actually there, annotate it as it comes. Include a TRANS comment with appropriate level if warranted.
2. If you must make changes, make the minimal change that is possible. If you can interpret the sentence by removing a single word (especially if it's a function word, determiner, etc.), put an X node on that word, and annotate the rest of the sentence as is (as if that X-ed word were not there).
3. Try to save the sentence, even if you have to take a larger piece of it out. If you can interpret the sentence by removing a phrase or several words that simply don't make sense, put an X node on that phrase, and annotate the rest of the sentence. Include a comment about the issue involved and include a TRANS comment with appropriate level.
4. Try not to represent OK English structure for something that is really not usable. If the sentence really is not interpretable or if there is too much word salad, use X or FRAG nodes at the top level (so that it doesn't look like a normal S). Include a comment about the issue involved, include a TRANS comment with appropriate level, and potentially save this sentence.

TRANS1 structures which are are unanalyzable or for which an analysis would represent non-English syntax.
TRANS2 missing major grammatical elements, such as auxiliaries or prepositions or other mistakes which
TRANS3 minor errors such as missing/extra determiners, incorrect verb tense, etc.

TOKEN: When there is incorrect tokenization (either in the source text or introduced in another layer of annotation), the comment TOKEN should be inserted in the comment field along with enough description to identify the location of the error easily.

TYPO: Typographical errors in the source text should be indicated with the comment TYPO in the comment field. The sentence should be annotated as though there were no

error.

3.10 Hyphenated Items

The following section is an adapted version of an email sent by Ann Taylor on December, 12, 2008.

3.10.1 Modifiers

There are two modifier labels: ADJP and NML

1) phrases labelled as ADJP
i) types with participles:

- noun/adv-participle

```
a (ADJP London/NNP - based/VBN)  company
a (ADJP so/RB - called/VBN) solution
(ADJP body/NN - piercing/VBG) armour
```

*but note:

```
a (NML tax/NN - raising/NN) platform
a (NML drug/NN - testing/NN) program
```

because the platform is not raising taxes nor the program testing drugs

- participle-particle

```
a (ADJP walled/VBN - in/RP) garden
a (ADJP drawn/VBN - out/RP) affair
```

ii) hyphenated adjectives with pseudo-participles (hard-hearted, single-minded, etc.) - label the pseudo-participle JJ and the first part by its POS (I use JJ where it's not clear what this is)

```
(ADJP hard/JJ - hearted/JJ)
(ADJP single/JJ - minded/JJ)
(ADJP honey/NN - tongued/JJ)
(ADJP white/JJ - haired/JJ)
```

iii) any modifier phrase that's *clearly* adjectival, no matter what it's structure

The test for this is that the *whole phrase* is gradable and (thus) can be modified by 'very'. Cases where the first part of the compound is gradable by itself are ambiguous (very high - level) and are treated as NML (not ADJP). Only use ADJP for these when the first part *isn't* gradable.

so:

```
(NML high/JJ - level/NN) talks
not *(ADJP high/JJ - level/NN) talks
```

but:

```
a (ADJP second/JJ - rate/NN) book
an (ADJP in/IN - depth/NN) interview
a (ADJP state/NN - of/IN - the/DT - art/NN) security system
a (ADJP run/VB - down/RP) section of town
a (ADJP right/JJ - wing/NN) view
```

It's especially convincing if the phrase can act as a predicate with 'very'

```
this technology is very state-of-the-art
the book is very second-rate
his views are very right-wing
```

The one exception to the above is *low key*, which should be ADJP despite its first part being gradable:

```
(ADJP low/JJ key/NN)
```

2) phrases labelled NML:

i) contain a nominal head (but cf. (iii) above)

```
(NML income tax) rates
```

ii) any modifier not an adjective or nominal; in practice PP or S (rarely ADVP?) In this case the NML dominates the PP or S and nothing else. Full structure of internal phrase is included.

```
(NP an
    (NML (PP on  (NP site)))
    camera)

(NP the
    (NML (S (NP-SBJ-1 no child)
```

62

```
                         (VP left)
                              (NP *-1)
                              (ADVP behind))))
          campaign)
```

3.10.2 Hyphenated Head Nouns

Label all parts as nominal (NN(S), etc.)

```
(NP the cease/NN - fire/NN)
(NP a real go/NN - getter/NN)
(NP in/NN - box/NN)

at (NP the get/NN - go/NN)
on (NP his say/NN - so/NN)
(NP a mother/NN - to/NN -be/NN)
```

3.10.3 The 'Pick-Up' Case

All verb+particle *modifiers* are labelled by parts and contained in an ADJP regardless
of whether they can also act as heads

```
(NP a (ADJP pick/VB up/RP) truck)
(NP a (ADJP stand/VB up/RP) guy)
```

When acting as heads, they are labeled according to B above.

```
(NP a pick/NN up/NN)
(NP two pick/NN ups/NNs)
```

When written as one word:

```
(NP a pickup/JJ truck)
(NP a pickup/NN)
```

This is consistent with the difference between:

```
(NP a (ADJP (NML 10/CD year/NN) old/JJ) child)
(NP two (NML 10/CD year/NN) olds/NNs)
```

3.10.4 Verbs

i) label parts by POS in isolation when clear

```
fine/RB - tuning/VBG
frog/NN - walked/VBN
spoon/NN - fed/VBN
right/RB - sizing/VBG
second/RB - guess/VB
two/CD - timing/VBG
campaign/NN -  managing/VBG
hand/NN - paints/VBZ
high/RB - fiving/VBG
```

ii) when POS is unclear, just use VB

```
pooh/VB - poohed/VBD
flip/VB - flopped/VBD
```

3.10.5 Affixes

Hyphenated affixes are not split. There is a list of what counts as an affix (see below). Basically if you would label it AFX if it was not hyphenated, it is considered an affix and not split.

If an affix is not hyphenated, label it AFX.

```
a (ADJP pro/AFX Saddam/NNP) rally
```

This approach makes for scope problems with things like

```
a (ADJP pro-George Bush) rally
```

The original idea was to label the pro-X bit JJ and the rest by POS,

(ADJP pro-George/JJ Bush/NNP) rally (ADJP pro-gay/JJ rights/NNS) legislation

but I'm currently leaning toward labelling all parts JJ (as part of a general drift towards labelling parts by the whole when the semantics isn't transparent)

(ADJP pro-gay/JJ rights/JJ) legislation

All names are split: Coca/NNP - Cola/NNP Al/NNP - Qaeda/NNP Wal/NNP - Mart/NNP

Split: Sino/JJ -Tibetan/JJ, Franco/JJ -Prussian/JJ

List of Unsplit Affixes

e- (e-mail, e-commerce)
a- (a-plenty)
u- u-turn
x- x-ray
anti-
agro- (agro-chemical)
be- be-dazzled
bi-
bio-
co-
counter- counter-terrorism, counter-insurgency
cross- cross-border, cross-country
cyber-
de-
eco-
-esque schula-esque
ex-
extra- extra-judicial
-fest Oz-fest
-fold four-fold
-gate
inter- inter-bank, inter-related
intra- intra-Arab
-itis niche-itis
-less lyric-less
macro-
mega-
micro-
mid- Mid-West, mid-seventies, mid-term, mid-2003
mini-
mm-hm interjection
mm-mm interjection
-most eastern-most
multi-
neo- neo-classical, neo-impressionism, Neo-Conservatives
non-
o-kay interjection
over- over-stimulate
pan-

para-
peri-
post-
pre-
pro-
pseudo-
quasi-
-rama
re-
semi-
sub-
super-
tri- tri-colored
uh-huh interjection
uh-oh interjection
ultra-
un- un-divine
uni- uni-polar
vice- vice-president
-wise points-wise