

TR-NLP-UT-2006-4*

GENIA Annotation Guidelines for Tokenization and POS Tagging

Yuka Teteisi², Jun'ichi Tsujii^{1,3,4}

¹University of Tokyo,

²Kogakuin University,

³University of Manchester

⁴National Centre for Text Mining

Date of Release: 15 November 2006

© Copyright 2006 Tsujii Laboratory, University of Tokyo

* All technical reports from Tsujii Laboratory can be found at
<http://www-tsujii.is.s.u-tokyo.ac.jp/TR/>

Introduction

This document describes the agreements between the annotators on the problematic cases in annotation of the GENIA corpus and changes to the original Penn Treebank POS annotation scheme[†]. In this document, examples are shown in boxes. A PTB-like format (where tokens are separated by a white space and each token may be followed by a slash followed by its part-of-speech) is used for describing examples in this document.

Overview of the scheme

The GENIA POS annotation scheme follows the Penn Treebank (PTB) POS annotation scheme. However, expressions not well-covered in the original PTB (e.g. technical terms, chemical formula, Linnean names) frequently appear in biomedical texts. As the POSs of the constituents can be determined independent of the context in which the term appears in most of such expressions, we chose to fix the POSs of the constituents of terms beforehand. Frequent technical terms in the GENIA corpus were compiled into a list where each constituent of each term in the list was assigned a POS. Some chemical expressions and abbreviations are compiled into regular expressions (e.g. IL-[0-9]/NN, CD[0-9]*-/JJ). Common expressions involving -ing and -ed forms of verbs (e.g. signaling/NN pathway/NN) and suffixes that determine the POS (e.g. -toxin/NN) are also added into the list. The expressions in the list used for annotation of GENIA version 3.x (called PREPOS hereafter) is added as Appendix 1 of this document.

The annotation process was as follows.

1. The text was tokenized using Penn tokenizer.
2. A series of perl scripts are run on the result to correct tokenization errors (mostly around chemical expressions like 2,3, 7,8-tetrachlorodibenzodioxin) and to assign POS to words by matching them with the expressions in PREPOS.
3. A modified version of JunK tagger[‡], which assigns only the POSs consistent with the pre-assigned POS, assigns the POS of remaining words.
4. Human annotators correct the result of the tagger. The pre-assigned POS are specially marked, so that in the human-correction phase, annotator can see which words are assigned POS based on the term list, meaning that they do not have to correct them (unless they are very confident of errors).

[†] Santorini, Beatrice. **Part-of-Speech Tagging Guidelines For The Penn Treebank Project (3rd Revision)**, Technical Report MS-CIS-90-47, University of Pennsylvania.

[‡] Kazama, Jun'ichi, Yusuke Miyao and Jun'ichi Tsujii. **A Maximum Entropy Tagger with Unsupervised Hidden Markov Models**. In the Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium. pp. 333-340, November 2001.

Tokenization

In GENIA 3.x and earlier versions, punctuations are separated from preceding tokens as are in PTB except in the names and numerical expressions. For example, the names and expressions shown below are treated as one token.

2,3,7,8-tetrachlorodibenzo-p-dioxin	
[3H]-Dexamethasone	
gamma(c)	
10(6)	("10 ⁶ ")
10(-10)-(-6)	(range; "from 10 ¹⁰ to 10 ⁶ ")
t(15;17)	(translocation)
(q11;q32)	(location in gene)
inv(14)(q11;q32)	(inversion at location in gene)
CD4(+)	("CD4-positive")
10(-4)/bp	(base pair)
5'	(terminus)
colombus:milk	(ratio)
1:3	(ratio)

If an abbreviation or acronym is parenthesized and put in between hyphenated words, it is separated from preceding and following words.

Interleukin (IL) -6

Parentheses and semicolons appearing as part of bibliographic information are separated from preceding and following words.

Immunity 1999 Mar ; 10 (3) : 399

Parentheses around the numbers (and other symbols) indicating references (e.g. [1]) are separated from the numbers.

[1]

Any inconsistent tokenizations of a word in the same abstract are left as they are unless the inconsistency occurs around hyphens, parenthesis and commas. For example, if "*HpaII*" and "*Hha I*" appear in one abstract, the inconsistency is not fixed.

In version 3.x and earlier, hyphenated words are not separated. This resulted in tokenization like "NF kappa-B-dependent" but such hyphenated expressions involving multi-word term are left as they are in order to keep the tokenization convention of the original PTB.

POS assignment

Prenominal Modifiers

Hyphenated words in the prenominal position are tagged as JJ as in the original PTB,

except those which are already known to be a substance name (and the like), which are tagged as NN. Hyphenated words that appear as nominal heads are tagged as NN.

granulocyte-macrophage/JJ lineage/NN
protein-protein/JJ interactions/NNS
gel-shift/JJ assay/NN
pro-B/JJ
mobility-shift/JJ assay/NN

A prenominal modifier where two or more words are concatenated by a slash is also tagged as JJ. Other tokens that are tagged as JJ in prenominal positions are shown below.

control (used in the meaning "of the control group")
human
adult
net (in "net weight" or "net effect")
immune (in "immune system")
all-trans
all-cis
"unilineage", "multilineage"
"*n*-fold" (where "n" stands for any given number: "*n*-fold" as a predicate, or complement of *be* is also tagged as JJ)
expressions of the form "XXX-" (or "XXX+") with the meaning "XXX negative" (or "XXX positive"). e.g. "CDnn+" (where "nn" stands for any given number, e.g. CD4+) modifying cell names such as "T cell"
Expressions of the form "Substance Name-/-", "Substance Name-/+ ", "Substance Name+/-" and "Substance Name+/+" modifying cell names (e.g. "Rel-/- T cells", "Rel+/+ T cells")
"5'" (or "5'-") and "3'" (or "3'-") (e.g. 5'/JJ flanking/JJ)
Expressions of the form "nn-unit" (where "nn" stands for any given number, e.g. nn-kD or nn-kilodalton) modifying nominal expressions
"helix-loop-helix"
"mean", meaning "average"
"few", "little", "former" and "latter" even when they appear as nominal heads.
"male" and "female" in prenominal positions
"IgH" in prenominal positions
non-Hodgkin/JJ 's/POS (disease)
postnominal modifier "upstream of". This expression can be paraphrased as "NN

which is upstream of . . . ". Hence, it is regarded as a reduced relative and tagged as JJ.
{situated,located}/JJ upstream/JJ of ...

terminal/JJ

tandem/JJ .

"split" in the prenominal position

prior/JJ to . . . (even if used adverbially)

On the other hand, the expressions shown below are tagged as NN in prenominal positions.

substance names, cell names, and their abbreviations (e.g. "IgM" in "Ig M receptor")

chemical elements and ions (e.g. Ca²⁺, Cl⁻)

"surface", meaning "on the surface of something"

"patient", meaning "person receiving medical treatment"

Some substance names are assigned JJ in prenominal positions and NN in NP head.

autocrine

endocrine

male

female

extra

promoter-enhancer

diploid

tetraploid

Expressions of the form "transXXX", "preXXX", "postXXX", "proXXX" where XXX is a name.

Expressions of the form "anti-XXX", where XXX is a name of a substance

"Ehrlichia chafeensis" is the Linnean name of an organism (bacteria) and is tagged as FW. Hence, "Anti-Ehrlichia" and "Anti-E" should be tagged as FW|JJ.

Expressions of the form "XXX-ant" (e.g. immunosuppressant, recombinant, mutant, variant) whose meaning is either "has an ability to XXX" or "something that XXXes".

Expressions of the form "XXX-ant {activated,inhibited}" modifying a nominal expression should be tagged as "XXX-ant/NN {activated,inhibited}/JJ".

Whether the "-ed" form of a verb should be tagged as JJ or VBN

It was not easy to determine whether an occurrence of a verb in its past participle ("-ed form") should be tagged as JJ or VBN. We used the following two criteria described in the PTB manual:

1. When a given past participle modifies a nominal expression and is preceded by a nominal expression modifying it or is hyphenated, it should be tagged as JJ.

10/CD genetic/JJ probe-based/JJ tests/NNS a/DT government-approved/JJ safety/NN seat/NN airline/NN leveraged/JJ buy-outs/NNS Thomson/NN consolidated/JJ earnings/NNS

2. When a given past participle is a complement of the be-verb or modifies a preceding nominal expression and is followed by a by-phrase, it should be tagged as VBN.

In addition to that, we used the following three rules:

1. If the past participle is followed by "state", "form", "pattern" or "fashion", it should be tagged as JJ.
2. Expressions of the form "nonXXX", "unXXX" or "inXXX" with the meaning "opposite of XXX" or "not XXX" should be tagged as JJ.
3. Expressions of the form "preXXX" (in which the prefix "pre" carries the meaning "before") should be tagged as JJ.

For some verbs we fixed the tag and added the verbs to the list of terms that have pre-fixed tags (see Appendix1, Appendix2).

The POS of a word that was not determined by the criteria above was determined by the following criteria:

1. Some words should be assigned different tags depending on their semantics. There are a class of verbs whose -ed forms are tagged as JJ when they refer to the condition of the body or the mind of a living organism and otherwise they are tagged as VBN (e.g. disturbed, depressed, crippled, inflamed).
2. The -ed forms that has no corresponding base verb form (e.g. threaded) should be tagged as JJ. The -ed forms of the verbs whose base forms are unambiguously intransitive should be tagged as JJ.
3. A past participle that refers to a position or place (e.g. located, positioned) should be tagged as JJ.
4. A past participle that refers to a result of a chemical reaction (e.g. phosphorylated, ionized; activated, inhibited) should be tagged as VBN.
- 5 The -ed forms with adjective entries in the third edition of Longman Dictionary of Contemporary English (LDOCE) should be tagged as JJ as long as the meanings are not significantly different from the dictionary definitions.

6. If the past participle refers to an event it should be tagged as VBN; if it refers to a state, it should be tagged as JJ. This criterion is from the original PTB Manual but it is overridden by the criteria 1-5 above.

7 The -ed forms with prefixes should generally be tagged in the same way as those without prefixes. (E.g.: coactivated and collocated should be tagged in the same way as activated and located, respectively; the former should generally be tagged as VBN and the latter should in general be tagged as JJ. Note, however, that there are some exceptions to this generalization, e.g., related/J J vs. correlated/VBN.)

8. If "XXXed NN" can be paraphrased as "NN that is (are) XXXed", "XXXed" should be tagged as VBN. If not, it should be tagged as JJ.

Whether the "-ing" form of a verb should be tagged as JJ, NN, or VBG

In the case of GENIA, the most problematic -ing word was "binding". Therefore, this section is divided into the subsection about "binding" and the one about other "-ing" forms.

Binding

As with other expressions, we fixed the POS of frequently appearing expressions involving the word "binding" (see Appendix1 and Appendix3). Those not in the lists are handled with the following criteria:

When "binding" is hyphenated, the hyphenated phrase should be tagged as JJ in the occurrences where it can be interpreted as a modifier and NN in the occurrence as the nominal head

the hormone-binding/JJ domain (HBD)

When "binding" is the head of the subject without any preceding determiner, possessive pronoun ("its"), adjective or adverb, it should be tagged as NN. When "binding" alone appears as an object of a verb by itself, it is tagged as NN | VBG .

Binding/NN is inhibited by an NFAT peptide spanning

When "binding" has a preceding determiner or adjective and is not the prenominal modifier, it should be tagged as NN.

FK506 had little effect on the binding/NN to the AP-1 site of PMA/ionomycin-induced nuclear factors, which were recognized with anti-JunD or c-Fos antibody.

When "binding" is immediately followed by a prepositional phrase headed by "of", it should be tagged as NN.

We observed binding/NN of STAT1, STAT3, and STAT4, but not STAT5 to all of these elements.

When "binding" functions as an object of an immediately preceding preposition, it should be usually tagged as NN|VBG. However, if "binding" appears in the expression "capable of binding", it is tagged as VBG.

We propose that HIV-1 tat may inhibit MIP-1 alpha expression by inducing MNP-1 expression in T-cells, probably by either competing with MNP-2 for binding/NN|VBG to the MIP-1 alpha promoter or by sequestering it into inactive forms

When "binding" should be immediately preceded by an NN(S), and if it is immediately followed by a to-phrase, it is tagged as NN|VBG. When "binding" is immediately preceded by an NN(S) but it is not followed by anything (i.e. if "NN binding" forms a phrase by itself), "binding" is tagged as NN.

Characterization of a new tissue-specific transcription factor binding/NN|VBG^ to the simian virus 40 enhancer TC-II (NF-kappa B) element.

When "binding" or the phrase headed by "binding" serves as an appositive to the noun immediately following the word "binding", it should be tagged as NN. (See also the following subsection).

-ing words other than "binding"

In general, when a present participle modifies a nominal expression and either it appears alone or is hyphenated, it should be tagged as JJ, and when "NN XXing" is a prenominal modifier or behaves as a nominal expression, "XXing" should be tagged as NN. (from PTB Manual). On the other hand, When the -ing form is in apposition to the following noun, it should be tagged as NN. In case of uncertainty the following criteria (listed in the order of preference) should be applied.

1. Those have entries in the PREPOS list (See Appendix 1) should be left as they are pre-tagged by preprocessors.
2. In some cases, the POS can be determined by the following noun. If the "-ing" form is followed by one of {activity, affinity, capacity, characteristics, effect, event, property, function, mechanism}, it is often an apposition, thus should be tagged as NN. However, it is not always so because "-ing" can be used to describe the nature of the activity, etc. Hence, be careful to consider the context when applying this rule.
3. The "-ing" forms before one of {form, pattern, fashion, state} should be tagged as JJ (the same as "-ed/JJ form", etc.)
4. The decision in the Appendix 3 should be followed if it does not contradict with the criteria 1-3 above.
5. The "-ing" forms with adjective entries in LDOCE should be tagged as JJ as long as

the meanings are not significantly different from the dictionary definitions.

6. Expressions of the form "preXXXing" in the prenominal position should be tagged as JJ (It is not likely that they appear as predicates).
7. If "NN1 XXing NN2" can be paraphrased as "NN1 that XXs NN2", "XXXing" should be tagged as VBG. (From PTB Manual)
8. If "XXing NN" can be paraphrased as "NN that XX(es) (something)", "XXXing" should be tagged as VBG. (From PTB Manual). Note that the criteria 7 and 8 are overridden by criteria 1-6.

Symbols

Symbols should be assigned tags according to their grammatical functions (in most cases, it should be tagged as NN or NNS.) and the use SYM should be avoided as much as possible.

Symbols to be tagged as NN(S)

Units composed of alphabets (e.g. nM, microM). The grammatical number should agree with the number of the numerator if the whole unit is in the form of a fraction.

Symbols not to be tagged as NN(S)

1. Roman numerals (I, II, III, ...) are tagged as CD.
2. List headers should be tagged as LS, even if they are numbers.

[/LBR 1/LS]/RBR

3. Field indicators (UI, TI, and AB) of the MEDLINE database should be tagged as LS.
4. "+", "-", and "+/-" used as conjunctions should be tagged as CC. The "x" (multiplication) sign should also be tagged as CC.
5. A combination of symbols to express a right arrow (e.g. "-->") should be tagged as TO.
6. The "=" (equal), "<" (less-than), and ">" (greater-than) signs: should be assigned the tag of the most appropriate paraphrase (e.g. "equal/JJ to", "is/VBZ equal to", "was/VBD equal to", "being/VBG equal to", "is/VBZ less than", "was/VBD greater than", "being/VBG less than").

Patients were classified as steroid sensitive (SS) if their morning FEV1 increased >/RBR 30%. However, it was still lower than that in healthy subjects (P < 0.01).

Proper nouns

Nominal expressions should be tagged as NN unless they are clearly proper names. More specifically, all nominal expressions should be tagged as NN except those listed below:

1. names of authors of abstracts
2. titles of journals, proceedings, and books. Plural forms should be tagged as

NNPS (e.g. "J./NNP Immunol./NNP Methods/NNPS").

3. initials of names of patients
4. country names
5. names of organizations (hospitals and laboratories)
6. names of months (January, February, . . .)

Note that names of diseases and the like that are taken from names of persons (e.g. "Down" in "Down syndrome") are tagged as NN. "Person Name-Person Name" is treated in the same way as NN-NN (e.g. "Epstein-Barr/JJ virus"). "Southern", "Northern", "Western" and "Northwestern" in the position of X in "X blotting" (experiment names) should be tagged as NN. It should be noted that words that begin with capital letters are not necessarily proper names (e.g. Copyright/NN 1997/CD Academic/NNP Press/NNP ./.).

Foreign words

In principle, the tag of a given foreign (Latin, Greek etc) word is determined by the following criteria:

1. Components of scientific (Linnean) names of living organisms should be tagged as FW (e.g. "Saccharomyces/FW cerevisiae/FW").
2. English equivalents of Greek letters should be tagged as NN (e.g. alpha/NN).
3. Components of idiomatic expressions originating from foreign languages that are frequently used in the literature of biology should be tagged as FW.

in/FW vitro/FW
in/FW vivo/FW
de/FW novo/FW
in/FW situ/FW

Other idioms of foreign origin are either tagged as FW or assigned tags that correspond to their functions according to the WSJ Corpus.

4. per/FW se/FW
5. et/FW al./FW
6. etc./FW i.e./FW (or, "i./FW e./FW")
7. e.g./FW (or, "e./FW g./FW")
8. ca./FW
9. vis-a-vis/FW
10. bona/FW fide/FW
11. via/IN
12. versus/CC
13. vs./CC

14. per/IN (except "per/FW se/FW")
15. plus
16. NN when used as a noun as in "a plus"
17. CC when used as a conjunction as in "A plus B"
18. vice/RB versa/RB .

. When an English word is hyphenated with a foreign word, the whole concatenated word is tagged as FW|XX where XX stands for the tag of the English word if the word is hyphenated with an English word (e.g. in/FW vivo-extended/FW|JJ).

Tags of other words of non-English origin are basically determined on word-by-word basis (See Appendix1).

Fragments around hyphens

In this case, the word beginning or ending with the hyphen (i.e. "XXX-" or "-YYY" in the above examples) should be tagged as if it were not prefixed or suffixed with the hyphen.

pre-/JJ and/CC post-transcription/NN
 pre-/JJ and/CC post-transcriptional/JJ
 pre-/RB and/CC post-transcriptionally/RB
 up-/RP and/CC down-regulates/VBZ
 interleukin/NN ((IL/NN)) -2/CD
 antigen-dependent/JJ and -independent/JJ
 nuclear/JJ factor/NN ((NF/NN)) -kappa/NN B/NN

However, "-XXXed" in expressions like: "Noun1/NN ((Abbreviation of Noun1/NN))-XXXed/?? Noun2/NN" should be tagged as JJ even if "-XXXed" is a word to be tagged as VBN by default. This is to keep the annotation of such expressions consistent with cases where the abbreviation enclosed by the parentheses does not appear ("Noun1-XXXed/JJ Noun2/NN").

lipopolysaccharide/NN ((LPS/NN)) -induced/JJ TF/NN expression/NN

Also, plural suffixes that stand alone as single words in expressions like "factor/NN ((s/NNS))" ("factor(s)" in the original text) should be tagged as NNS. Third person singular present suffixes should be tagged as VBZ as in "play/VBP ((s/VBZ))".

Acronyms and abbreviations

The POS of an acronym or an abbreviation should conform to the POS of the original form spelt out in the same abstract if there is any. For example, "FasL" in an abstract titled "Regulation of fas-ligand expression", should be tagged as JJ because "fas-ligand" is a hyphenated prenominal modifier and is tagged as JJ and "FasL" is the abbreviation

of "fas-ligand".

Acronyms whose original forms cannot be found in the same abstract should by default be tagged as NN. Acronyms should in principle be tagged consistently within a single abstract. Hence, the same acronym might be assigned different tags in different abstracts.

Whether an acronym or an abbreviation should be NN or NNS

If an acronym or an abbreviation retains the original plural suffix it should be tagged as NNS.

EBVs/NNS	(Epstein-Barr viruses)
hrs/NNS	(hours)

Acronyms and abbreviations which do not have plural suffixes generally fall under one of the categories listed below and should be tagged accordingly.

1. The grammatical number of acronyms should be made consistent with the original forms that are spelt out in the same abstract. Acronyms that are tagged as NNS in other parts of an abstract should also be tagged as NNS in the prenominal position.

protein/NN kinase/NN C/NN (/ (PKC/NN)/)

If an acronym whose original form is plural appears both with and without the plural suffix, those with the suffix should be tagged as NNS and those without the suffix should be tagged as NN. Hence, something like the following could happen against the principle above in one and the same abstract:

Epstein-Barr viruses (EBV/NN) ... EBVs/NNS ...
--

Note that, in this case, if the form "EBVs" did not appear, "EBV" would be tagged as NNS throughout the abstract.

2. If the same acronym appears in both positions that are restricted to singular forms and positions that are restricted to plural forms, tokens in each environment should be tagged as NN and NNS, respectively. Those that appear in positions where both singular and plural forms can appear should be tagged as NN. In this case, if an instance of the acronym immediately following the original form enclosed by parentheses doesn't have the plural suffix, it should be tagged as NNS. (The acronym in this position is considered to be an abbreviation of the plural form.)

3. In contrast to acronyms, abbreviations of units which have lost their original plural suffixes should simply be tagged as NN even if their original forms can be clearly detected to be plural.

base pairs (bp/NN)

Even when preceded by numbers larger than 1, abbreviations of units that do not have

plural suffixes should be tagged as NN (e.g. 2/CD hr/NN).

4. "pp." which abbreviates for "pages" should be tagged as NNS.

Grammatical numbers in general cases

Words that end in "s" whose singular and plural forms are identical should be tagged as NN by default. If they appear in positions where only plural forms can appear, they should be tagged as NNS.

E.g.:

by means/NN of (Although the WSJ Corpus tag this as NNS, we tag it as . NN.)

{a, the} series/NN

by all means/NNS

The grammatical number of fractions should agree with the number of the numerator (e.g. sites/cell/NNS, fmol/micrograms/NN^s).

Slashes

Tentatively, slashes should not be separated from preceding and following words except cases in which it is explicitly stated that they should stand alone as independent words.

Words concatenated by slashes should be tagged according to the following guidelines:

Expressions in which two words whose independent POSes are either NN or JJ are concatenated and which function as prenominal modifiers should be tagged as JJ. (In such cases, the function of slashes are almost identical to that of hyphens.)

Other words concatenated by slashes are in general of the same part of speech. Hence, the whole concatenated unit should be assigned the POS common to both components.

The treatment of slashes will be examined later and the segmentation and tags of words containing slashes will be fixed according to the principles that will be established then.

Errors in the original text

Spelling errors in the original abstract should be left as they are and tagged as if they are correctly spelt.

Alcoholic/JJ extracts/NNS prepared/VBN form/IN Arnicae/FW flos/FW

In this case, it can be inferred from the context that "form" is a misspelling of "from". Hence, it should be tagged as "form/IN" as above. Note that this rule goes against the POS-tagging convention stated in the PTB Manual.

In case of disagreement between subject and verb, the POS of the subject should agree with the grammatical number of the verb.

The/DT results/NN indicates/VBZ that/IN carboranes/NNS are/VBP applicable/JJ as/IN the/DT hydrophobic/JJ moiety/NN of/IN biologically/RB active/JJ molecules/NNS ./.
--

Miscellaneous

1. "ago" in expressions like "several year s ago" should be tagged as RB.
2. "irrespective of ..." should be tagged as "irrespective/RB of/IN ..." since it modifies the whole sentence adverbially.
3. "post" used in the meaning of "after" should be tagged as IN.
4. All instances of "when" should be tagged as WRB. Though the PTB Manual states that it should be tagged as IN if it is used "in the sense of `if' ", we don't adopt this rule since it is often difficult to distinguish cases in which "when" should be tagged as IN from cases in which it should be tagged as WRB.

Acknowledgment

The authors are grateful to annotators (Yusuke Kubota, Kyoko Takano, Camille Yamada, and Emi Maekawa) and colleagues (Tomoko Ohta, Duo Wang, Yusuke Miyao and Akane Yakushiji) that helped the construction of the corpus.

Appendix 1

PREPOS data

The following expressions are technical terms and other expressions that are assigned pre-fixed POSs independent of the context they appear by "prepos.pl" before mechanical POS-tagging.

Foreign Terms

Linnean Names of Organisms(treated as Latin)

B. burgdorferi	=> B./FW burgdorferi/FW
Borrelia burgdorferi	=> Borrelia/FW burgdorferi/FW
C. Sub.	=> C./JJ Sub./NN
C. albicans	=> C./FW albicans/FW
C. neoformans	=> C./FW neoformans/FW
Candida albicans	=> Candida/FW albicans/FW
Catalytic Subunit	=> Catalytic/JJ Subunit/NN
Cryptococcus neoformans	=> Cryptococcus/FW neoformans/FW
D. melanogaster	=> D./FW melanogaster/FW
Drosophila melanogaster	=> Drosophila/FW melanogaster/FW
E. coli	=> E./FW coli/FW
Escherichia coli	=> Escherichia/FW coli/FW
L. donovani	=> L./FW donovani/FW
L. monocytogenes	=> L./FW monocytogenes/FW
L. pneumophila	=> L./FW pneumophila/FW
Legionella pneumophila	=> Legionella/FW pneumophila/FW
Leishmania donovani	=> Leishmania/FW donovani/FW
Listeria monocytogenes	=> Listeria/FW monocytogenes/FW
P. aeruginosa	=> P./FW aeruginosa/FW
P. falciparum	=> P./FW falciparum/FW
Pseudomonas aeruginosa	=> Pseudomonas/FW aeruginosa/FW

Plasmodium falciparum => Plasmodium/FW falciparum/FW
 S. aureus => S./FW aureus/FW
 S. cerevisiae => S./FW cerevisiae/FW
 S. oedipus => S./FW oedipus/FW
 S. typhimurium => S./FW typhimurium/FW
 Saccharomyces cerevisiae => Saccharomyces/FW cerevisiae/FW
 Saguinus oedipus => Saguinus/FW oedipus/FW
 Salmonella typhimurium => Salmonella/FW typhimurium/FW
 Staphylococcus aureus => Staphylococcus/FW aureus/FW
 T. annulata => T./FW annulata/FW
 T. pallidum => T./FW pallidum/FW
 Theileria annulata => Theileria/FW annulata/FW
 Treponema pallidum => Treponema/FW pallidum/FW

Other Terms of Latin Origin

bona fide => bona/FW fide/FW
 in vitro => in/FW vitro/FW
 In vitro => In/FW vitro/FW
 in vivo => in/FW vivo/FW
 In vivo => In/FW vivo/FW
 ex vivo => ex/FW vivo/FW
 Ex vivo => Ex/FW vivo/FW
 de novo => de/FW novo/FW
 De novo => De/FW novo/FW
 in situ => in/FW situ/FW
 In situ => In/FW situ/FW
 in-situ => in-situ/FW
 In-situ => In-situ/FW
 per se => per/FW se/FW
 et al => et/FW al/FW
 et al. => et/FW al./FW

etc => etc/FW
etc. => etc./FW
ie. => i.e./FW
i.e. => i.e./FW
i. e. => i./FW e./FW
e.g. => e.g./FW
e. g. => e./FW g./FW
ca. => ca./FW
versus. => versus/CC
via => via/IN
vs. => vs./CC
per => per/IN

Greek Letters

alpha => alpha/NN
beta => beta/NN
gamma => gamma/NN
kappa => kappa/NN
Alpha => Alpha/NN
Beta => Beta/NN
Gamma => Gamma/NN
Kappa => Kappa/NN

Other terms of foreign origin

Liuwei Dihuang Pills => Liuwei/FW Dihuang/FW Pills/NNS
bacille Calmette-Guerin vaccine => bacille/FW Calmette-Guerin/FW vaccine/NN

Confusing Singular/Plural Forms

anlage => anlage/NN
foci => foci/NNS

focus => focus/NN
mice => mice/NNS
nuclei => nuclei/NNS
nucleus => nucleus/NN
pancreas => pancreas/NN
pancreata => pancreata/NNS
sera => sera/NNS
serum => serum/NN
women => women/NNS

Multi-word prepositions and other fixed expressions

According to => According/VBG to/IN
As well as => As/RB well/RB as/IN
Due to => Due/JJ to/TO
In general => In/IN general/JJ
In particular => In/IN particular/JJ
Instead of => Instead/RB of/IN
That is, => That/DT is/VBZ ,/,
Whereas => Whereas/IN
a means => a/DT means/NN
according to => according/VBG to/IN
as well => as/RB well/RB
as well as => as/RB well/RB as/IN
by means of => by/IN means/NN of/IN
due to => due/JJ to/TO
each other => each/DT other/JJ
in general => in/IN general/JJ
in particular => in/IN particular/JJ
instead of => instead/RB of/IN
so called => so/RB called/VBN
such as => such/JJ as/IN

that is, => that/DT is/VBZ ,/,
whereas => whereas/IN
at least => at/IN least/JJS

Numeric Expressions

Units: treat as nouns

Tag prenominal "number-unit" (number hyphenated with unit) expressions as adjectives.

microM => microM/NN
fmol/10(7) => fmol/10(7)/NN
bp => bp/NN

Symbols

[=> [(
] =>]/
+/- => +/-/CC
+ => +/CC
- => -/CC
> => >/JJR (May be RBR, depending on context)
< => </JJR (May be RBR, depending on context)
= => =/JJ (May be RB, depending on context)

Phrases Treated as Adjectives

Note:

- For gerunds and participles of verbs, refer to Appendix 2 and 3.
- Treat as a noun when it is a nominal head.

Abbreviations

CD1+ReIB+ => CD1+ReIB+/JJ

CD14(+) => CD14(+)/JJ
CD14+ => CD14+/JJ
CD34(+) => CD34(+)/JJ
CD4(+) => CD4(+)/JJ
CD4(-) => CD4(-)/JJ
CD4+ => CD4+/JJ
CD4- => CD4-/JJ
CD68(+) => CD68(+)/JJ
CD68(-) => CD68(-)/JJ
CD68+ => CD68+/JJ
CD68- => CD68-/JJ
CD8(+) => CD8(+)/JJ
CD8(-) => CD8(-)/JJ
CD8+ => CD8+/JJ
CD8- => CD8-/JJ
FCS+ => FCS+/JJ
FCS- => FCS-/JJ
GD3+ => GD3+/JJ
GD3- => GD3-/JJ
HTLV-I(+) => HTLV-I(+)/JJ
p53-/- => p53-/-/JJ

Other Adjectives

(AOP)-RANTES-triggered => (AOP)-RANTES-triggered/JJ
3' => 3'/JJ
3-hydroxy-3-methylglutaryl => 3-hydroxy-3-methylglutaryl/JJ
5' => 5'/JJ
5-Isoquinolinyl => 5-Isoquinolinyl/JJ
A-independent => A-independent/JJ
B-lymphoblastoid => B-lymphoblastoid/JJ
Ca(2+)-dependent => Ca(2+)-dependent/JJ

FCS	=> FCS/JJ
Human	=> Human/JJ
Molecular	=> Molecular/JJ
PKC-dependent	=> PKC-dependent/JJ
Stat1-dependent	=> Stat1-dependent/JJ
T-lymphoblastoid	=> T-lymphoblastoid/JJ
adult	=> adult/JJ
anti-Ig	=> anti-Ig/JJ NN
bHLH	=> bHLH/JJ
basal	=> basal/JJ
bovine	=> bovine/JJ
chromosomal	=> chromosomal/JJ
chronic	=> chronic/JJ
cis	=> cis/JJ
dependent	=> dependent/JJ
dibutryl	=> dibutryl/JJ
dificient	=> dificient/JJ
dominant-negative	=> dominant-negative/JJ
elevated	=> elevated/JJ
endocrine	=> endocrine/JJ
erythroid	=> erythroid/JJ
granulocyte-macrophage	=> granulocyte-macrophage/JJ
helix-loop-helix	=> helix-loop-helix/JJ
heterozygous	=> heterozygous/JJ
heterozygous	=> heterozygous/JJ
homozygous	=> homozygous/JJ
human	=> human/JJ

◆anti- : depends on context, if prenominal it's JJ but if it's nominal head it's NN

◆basic helix-loop-helix

immune	=> immune/JJ
inflammatory	=> inflammatory/JJ
kinase-dependent	=> kinase-dependent/JJ
located	=> located/JJ
lymphoid	=> lymphoid/JJ
mature	=> mature/JJ
mononuclear	=> mononuclear/JJ
murine	=> murine/JJ
nevoid	=> nevoid/JJ
novel	=> novel/JJ
novel	=> novel/JJ
parental	=> parental/JJ
present	=> present/JJ
present	=> present/JJ
related	=> related/JJ
related	=> related/JJ
resistant	=> resistant/JJ
rheumatoid	=> rheumatoid/JJ
sIgD-crosslinking	=> sIgD-crosslinking/JJ
sIgM-crosslinking	=> sIgM-crosslinking/JJ
semimature	=> semimature/JJ
synovial	=> synovial/JJ
terminal	=> terminal/JJ
terminal	=> terminal/JJ
tissue-specific	=> tissue-specific/JJ
tolerant	=> tolerant/JJ
tonsillar	=> tonsillar/JJ
trans-acting	=> trans-acting/JJ
trans	=> trans/JJ
transcription-independent/JJ	=> transcription-independent/JJ
transient	=> transient/JJ

umbilical	=> umbilical/JJ
unilineage	=> unilineage/JJ
zinc-finger	=> zinc-finger/JJ

Verbs

evaginate => evaginate/VBP|VB

Names with Hyphens: treated as nouns

(AOP)-RANTES	=>	(AOP)-RANTES/NN
12-myristate	=>	12-myristate/NN
13-acetate	=>	13-acetate/NN
20-Epi	=>	20-Epi/NN
20-epi	=>	20-epi/NN
25-Dihydroxyvitamin	=>	25-Dihydroxyvitamin/NN
3'-UTR	=>	3'-UTR/NN
32Dcl3	=>	32Dcl3/NN
6TG1.1	=>	6TG1.1/NN
AIR-1	=>	AIR-1/NN
Academic Press	=>	Academic/NNP Press/NNP
B*CaM-K	=>	B*CaM-K/NN
BCL-2	=>	BCL-2/NN
CDR2-loop	=>	CDR2-loop/NN
CTLA-4	=>	CTLA-4/NN
D1/CDR2-loop	=>	D1/CDR2-loop/NN
D1/CDR3-loop	=>	D1/CDR3-loop/NN
FGF-1	=>	FGF-1/NN
Fli-1	=>	Fli-1/NN
G-CSF	=>	G-CSF/NN
GSK-3beta	=>	GSK-3beta/NN
Glucocorticoid	=>	Glucocorticoid/NN

HCD57-SREI	=>	HCD57-SREI/NN	◆name of a cell line
HES-1	=>	HES-1/NN	
HL60	=>	HL60/NN	
HS-72	=>	HS-72/NN	
HT-2	=>	HT-2/NN	
HUVECs	=>	HUVECs/NNS	
Hepa-1	=>	Hepa-1/NN	
ICAM-1	=>	ICAM-1/NN	
IFNs	=>	IFNs/NNS	
IL-12R	=>	IL-12R/NN	
IL-1beta	=>	IL-1beta/NN	
IL-1beta	=>	IL-1beta/NN	
IL-2R	=>	IL-2R/NN	
IL-	=>	IL-/NN	
JNK	=>	JNK/NN	
LEF-1	=>	LEF-1/NN	
LEF-1	=>	LEF-1/NN	
LMP-1	=>	LMP-1/NN	
LP	=>	LP/NN	
M4	=>	M4/NN	
MAPK	=>	MAPK/NN	
MCP-1	=>	MCP-1/NN	
ML-1	=>	ML-1/NN	
MO7e	=>	MO7e/NN	◆name of a cell line
NAD(P)H:quione	=>	NAD(P)H:quione/NN	
NF-AT-	=>	NF-AT-/NN	
NF-AT _c	=>	NF-AT _c /NN	
NF-AT _c	=>	NF-AT _c /NN	
NK	=>	NK/NN	

NSAIDs	=>	NSAIDs/NNS
Notch1/TAN-1	=>	Notch1/TAN-1/NN
OH	=>	OH/NN
PGG-Glucan	=>	PGG-Glucan/NN
SHP-1	=>	SHP-1/NN
SOCS-3	=>	SOCS-3/NN
STAT-5	=>	STAT-5/NN
STAT5a	=>	STAT5a/NN
Spi-1	=>	Spi-1/NN
T	=>	T/NN
TNF	=>	TNF/NN
TNF	=>	TNF/NN
U-937	=>	U-937/NN
UF-1	=>	UF-1/NN
VCAM-1	=>	VCAM-1/NN
VCAM-1	=>	VCAM-1/NN
X-chromosome	=>	X-chromosome/NN
activin	=>	activin/NN
aminooxypentane	=>	aminooxypentane/NN
anorexia	=>	anorexia/NN
antigen	=>	antigen/NN
bcl-xL	=>	bcl-xL/NN
beta-catenin	=>	beta-catenin/NN
beta-globin	=>	beta-globin/NN
beta-hydroxysteroid	=>	beta-hydroxysteroid/NN
breast	=>	breast/NN
c-Myc	=>	c-Myc/NN
c-myc	=>	c-myc/NN
cytochalasin	=>	cytochalasin/NN
dehydrastase/dimerization	=>	dehydrastase/dimerization/NN
elatase	=>	elatase/NN

factor-1	=>	factor-1/NN	
fibroblast	=>	fibroblast/NN	
fos	=>	fos/NN	
glycosphingolipid	=>	glycosphingolipid/NN	
hIL-5	=>	hIL-5/NN	
haemoglobin	=>	haemoglobin/NN	
hybrid	=>	hybrid/NN	
hydroxysteroid	=>	hydroxysteroid/NN	
immunoreceptor	=>	immunoreceptor/NN	
interleukin-10	=>	interleukin-10/NN	
interleukin-1	=>	interleukin-1/NN	
interleukin-3	=>	interleukin-3/NN	
interleukin	=>	interleukin/NN	
isoforms	=>	isoforms/NNS	(sometimes VBZ)
lipid	=>	lipid/NN	
macrophage	=>	macrophage/NN	
mg-1	=>	mg-1/NN	
mispairing	=>	mispairing/NN	
molecule-1	=>	molecule-1/NN	
nucleotide	=>	nucleotide/NN	
p21	=>	p21/NN	
p27	=>	p27/NN	
p45	=>	p45/NN	
p62	=>	p62/NN	
p70(S6)-kinase	=>	p70(S6)-kinase/NN	
p70	=>	p70/NN	
pRb	=>	pRb/NN	
phi	=>	phi/NN	
prednisolone	=>	prednisolone/NN	
pro-interleukin	=>	pro-interleukin/NN	

proIL-1beta	=>	proIL-1beta/NN	
proteasome	=>	proteasome/NN	
purine-box/antigen	=>	purine-box/antigen/NN	
rapamycin	=>	rapamycin/NN	
ras	=>	ras/NN	
s6k	=>	s6k/NN	
s	=>	s/NNS	
sera	=>	sera/NNS	(plural of serum)
serine/threonine	=>	serine/threonine/NN	
signal	=>	signal/NN	
stimulus	=>	stimulus/NN	
thiol/disulfide	=>	thiol/disulfide/NN	
thrombin	=>	thrombin/NN	
tumor	=>	tumor/NN	
1 alpha,25(OH)2D3	=>	1/CD alpha,25(OH)2D3/NN	
1 alpha,25-(OH)2D3	=>	1/CD alpha,25-(OH)2D3/NN	
1 alpha,25-dihydroxyvitamin D3	=>	1/CD alpha,25-dihydroxyvitamin/NN D3/NN	
1,25 dihydroxyvitamin D3	=>	1,25/CD dihydroxyvitamin/NN D3/NN	
1,25(OH)2D3	=>	1,25(OH)2D3/NN	
1,25-(OH)2D3	=>	1,25-(OH)2D3/NN	
1,25-dihydroxycholecalciferol	=>	1,25-dihydroxycholecalciferol/ NN	
1,25-dihydroxyvitamin D3	=>	1,25-/CD dihydroxyvitamin/NN D3/NN	
1,25-dihydroxyvitamin D3	=>	1,25-dihydroxyvitamin/NN D3/NN	
2,3,7,8-tetrachlorodibenzo-p-dioxin	=>	2,3,7,8-tetrachlorodibenzo-p-di oxin/NN	

3-kinase	=>	3-kinase/NN	
5'-(N-ethyl)-carboxamido adenosine	=>	5'-(N-ethyl)-carboxamido adenosine/NN	
5'-flanking region	=>	5'-flanking/JJ region/NN	
ADP-ribose	=>	ADP-ribose/NN	
AIDS	=>	AIDS/NN	
AIR-1	=>	AIR-1/NN	
AP-1 element	=>	AP-1/NN element/NN	
AP-1 site	=>	AP-1/NN site/NN	
AP-1	=>	AP-1/NN	name of transcription factor
AP-2	=>	AP-2/NN	name of transcription factor
AP1	=>	AP1/NN	◆ name of transcription factor
APL	=>	APL/NN	◆ leukemia
AT2.5	=>	AT2.5/NN	◆ cell line
ATF-1	=>	ATF-1/NN	
ATRA	=>	ATRA/NN	◆ all-trans retinoic acid
Ab	=>	Ab/NN	◆ antibody
Abs	=>	Abs/NNS	◆ antibodies
Ah-receptor	=>	Ah-receptor/NN	
Ah	=>	Ah/NN	
B cell	=>	B/NN cell/NN	
B cells	=>	B/NN cells/NNS	
B lymphocyte	=>	B/NN lymphocyte/NN	
B lymphocytes	=>	B/NN lymphocytes/NNS	
B lymphoma	=>	B/NN lymphoma/NN	

B*CaM-K	=>	B*CaM-K/NN	
B-cell	=>	B-cell/NN	
B-cells	=>	B-cells/NNS	
B104	=>	B104/NN	◆cell line
B95-8	=>	B95-8/NN	
BALB/c 3T3	=>	BALB/c/NN 3T3/NN	◆cell line
BCL-6	=>	BCL-6/NN	
BL2	=>	BL2/NN	◆cell line
BSAP	=>	BSAP/NN	
BW5147	=>	BW5147/NN	◆cell line
BZLF1 promoter	=>	BZLF1/NN promoter/NN	
BZLF1	=>	BZLF1/NN	
Bcl-2	=>	Bcl-2/NN	
CAT gene	=>	CAT/NN gene/NN	
CD14	=>	CD14/NN	
CD40	=>	CD40/NN	
CD4	=>	CD4/NN	
CD8	=>	CD8/NN	
CSF-1	=>	CSF-1/NN	
Ca(2+)-ATPase	=>	Ca(2+)-ATPase/NN	
Ca(2+)-ionophore	=>	Ca(2+)-ionophore/NN	
Ca2+	=>	Ca2+/NN	
Cp	=>	Cp/NN	
CsA	=>	CsA/NN	
DM	=>	DM/NN	
DN	=>	DN/JJ	◆dominant-neg ative
DNA binding	=>	DNA/NN binding/NN	
DNA sequences	=>	DNA/NN sequences/NNS	
DNA		DNA/ NN	

DNase-I	=>	DNase-I/NN	
E(gre)	=>	E(gre)/NN	
E1A	=>	E1A/NN	
E2F-1	=>	E2F-1/NN	
E2F	=>	E2F/NN	
EBNA-2	=>	EBNA-2/NN	
EBNA2	=>	EBNA2/NN	
EBV	=>	EBV/NN	◆Epstein-Barr Virus
EGR-1	=>	EGR-1/NN	
ER	=>	ER/NN	
Egr-1	=>	Egr-1/NN	
Elf-1	=>	Elf-1/NN	
Epstein-Barr virus	=>	Epstein-Barr/JJ virus/NN	◆Epstein, Barr : person
FDC	=>	FDC/NN	
FK506	=>	FK506/NN	
Fos	=>	Fos/NN	
G(Anh)MTetra	=>	G(Anh)MTetra/NN	
GATA-1 gene	=>	GATA-1/NN gene/NN	
GATA-1	=>	GATA-1/NN	
GC box	=>	GC/NN box/NN	
GCR	=>	GCR/NN	
GD3	=>	GD3/NN	
GM-CSF gene	=>	GM-CSF/NN gene/NN	
GM-CSF	=>	GM-CSF/NN	
GP+E-86	=>	GP+E-86/NN	◆cell line
GR mRNA	=>	GR/NN mRNA/NN	
GR	=>	GR/NN	
H2O2	=>	H2O2/NN	
HB24	=>	HB24/NN	

HCD57-SREI	=>	HCD57-SREI/NN	
HES-1	=>	HES-1	
HIV enhancer	=>	HIV/NN enhancer/NN	
HIV long terminal repeat	=>	HIV/NN long/JJ terminal/JJ repeat/NN	
HIV-1 LTR	=>	HIV-1/NN LTR/NN	◆human
HIV-1 LTR	=>	HIV-1/NN LTR/NN	immunodeficient virus long term repeat
HIV-1	=>	HIV-1/NN	◆human immunodeficient virus I
HIV-TF1	=>	HIV-TF1/NN	
HIV	=>	HIV/NN	◆human immunodeficient virus
HL-60	=>	HL-60/NN	◆cell line
HLA-DR and -DP	=>	HLA-DR/NN and/CC -DP/NN	
HLA-DRA and DQB	=>	HLA-DRA/NN and/CC DQB/NN	
HPB.ALL	=>	HPB.ALL/NN	◆cell line
HTLV-I	=>	HTLV-I/NN	◆human T-cell leukemia virus
HUT78	=>	HUT78/NN	◆cell line
HeLa	=>	HeLa/NN	◆cell line
Hepa-1	=>	Hepa-1/NN	
I epsilon exon	=>	I/NN epsilon/NN exon/NN	
I kappa B alpha	=>	I/NN kappa/NN B/NN alpha/NN	
I kappa B-alpha	=>	I/NN kappa/NN B-alpha/NN	
I kappa B	=>	I/NN kappa/NN B/NN	

ICAM-1	=>	ICAM-1/NN	
ICAM-1	=>	ICAM-1/NN	
IFN alpha	=>	IFN/NN alpha/NN	
IFN-MCP	=>	IFN-MCP/NN	
IFN-alpha	=>	IFN-alpha/NN	
IFN-gamma	=>	IFN-gamma/NN	
IL-1 beta	=>	IL-1/NN beta/NN	◆interleukin
IL-10	=>	IL-10/NN	◆interleukin
IL-12	=>	IL-12/NN	◆interleukin
IL-13R	=>	IL-13R/NN	
IL-1	=>	IL-1/NN	
IL-2 gene	=>	IL-2/NN gene/NN	
IL-2 gene	=>	IL-2/NN gene/NN	
IL-2 promoter	=>	IL-2/NN promoter/NN	
IL-2 promoter	=>	IL-2/NN promoter/NN	◆interleukin
IL-2	=>	IL-2/NN	◆interleukin
IL-2R alpha	=>	IL-2R/NN alpha/NN	◆interleukin
IL-2R	=>	IL-2R/NN	
IL-3	=>	IL-3/NN	
IL-4 gene	=>	IL-4/NN gene/NN	
IL-4	=>	IL-4/NN	◆interleukin
IL-4R	=>	IL-4R/NN	
IL-5	=>	IL-5/NN	
IL-5	=>	IL-5/NN	
IL-5	=>	IL-5/NN	◆interleukin
IL-6 gene	=>	IL-6/NN gene/NN	
IL-6	=>	IL-6/NN	◆interleukin
IL-8	=>	IL-8/NN	◆interleukin
IRF-1	=>	IRF-1/NN	
Id3	=>	Id3/NN	
Ig heavy chain mu E1 site	=>	Ig/NN heavy/JJ chain/NN	

		mu/NN E1/NN site/NN	
IgD	=>	IgD/NN	
IgE	=>	IgE/NN	
IgH	=>	IgH/NN	
IgM	=>	IgM/NN	
J.Jhan	=>	J.Jhan/NN	◆cell line
JAK3	=>	JAK3/NN	
JNKs	=>	JNKs/NNS	
Jun	=>	Jun/NN	
JunD	=>	JunD/NN	
Jurkat	=>	Jurkat/NN	◆cell line
K-562	=>	K-562/NN	◆cell line
K562	=>	K562/NN	◆cell line
LAK	=>	LAK/NN	
LFA-1	=>	LFA-1/NN	
LMP1(1-231)	=>	LMP1(1-231)/NN	
LMP1	=>	LMP1/NN	
LPS	=>	LPS/NN	
LTB4	=>	LTB4/NN	
LTR	=>	LTR/NN	
LTR	=>	LTR/NN	
LYSP100	=>	LYSP100/NN	
M-CSF	=>	M-CSF/NN	
M-TAT	=>	M-TAT/NN	
M1	=>	M1/NN	◆cell line
M4	=>	M4/NN	
MCP-1	=>	MCP-1/NN	
MEL	=>	MEL/NN	
MO7e	=>	MO7e/NN	
MR	=>	MR/NN	
NAC	=>	NAC/NN	

NB4	=>	NB4/NN	◆cell line
NF kappa B	=>	NF/NN kappa/NN B/NN	
NF-AT	=>	NF-AT/NN	
NF-ATp	=>	NF-ATp/NN	
NF-kB	=>	NF-kB/NN	
NF-kappa B p65	=>	NF-kappa/NN B/NN p65/NN	
NF-kappa B site	=>	NF-kappa/NN B/NN site/NN	
NF-kappa B-dependent genes	=>	NF-kappa/NN B-dependent/JJ genes/NNS	
NF-kappa B	=>	NF-kappa/NN B/NN	
NF-kappaB	=>	NF-kappaB/NN	
NFAT-1	=>	NFAT-1/NN	
NFAT1	=>	NFAT1/NN	
NFAT	=>	NFAT/NN	
NFATp	=>	NFATp/NN	
NK	=>	NK/NN	
NS-Meg	=>	NS-Meg/NN	
Na(+)-H(+)-antiport	=>	Na(+)-H(+)-antiport/NN	
Na(+)-H+ antiport	=>	Na(+)-H+/JJ NN antiport/NN	
Nef	=>	Nef/NN	
Northern blot	=>	Northern/NN blot/NN	
Northern blotting	=>	Northern/NN blotting/NN	
OBF-1	=>	OBF-1/NN	
OCA-B	=>	OCA-B/NN	
OCT-1	=>	OCT-1/NN	
Oct-1	=>	Oct-1/NN	
Oct-2	=>	Oct-2/NN	
P3HR-1	=>	P3HR-1/NN	
PBL	=>	PBL/NN	
PGE2	=>	PGE2/NN	
PHA	=>	PHA/NN	

PIH	=>	PIH/NN	
PKC-zeta	=>	PKC-zeta/NN	
PKC	=>	PKC/NN	
PMA	=>	PMA/NN	
PML/RARalpha	=>	PML/RARalpha/NN	
PML	=>	PML/NN	
PU.1	=>	PU.1/NN	
R24	=>	R24/NN	
RA	=>	RA/NN	
Raji	=>	Raji/NN	◆cell line
Rb	=>	Rb/NN	
RelA	=>	RelA/NN	
RelB	=>	RelB/NN	
S phase	=>	S/NN phase	
SK-N-Be(2)	=>	SK-N-Be(2)/NN	◆cell line
SLP-76	=>	SLP-76/NN	
SP1 binding site	=>	SP1/NN binding/NN site/NN	
SRY	=>	SRY/NN	
STAT1 alpha	=>	STAT1/NN alpha/NN	
STAT3	=>	STAT3/NN	
STAT4	=>	STAT4/NN	
STAT5	=>	STAT5/NN	
SV40	=>	SV40/NN	
Southern blotting	=>	Southern/NN blotting/NN	
Southern blot	=>	Southern/NN blot/NN	
Southwestern blot	=>	Southwestern/NN blot/NN	
Southwestern blotting	=>	Southwestern/NN blotting/NN	
Sp1 site	=>	Sp1/NN site/NN	
Sp1	=>	Sp1/NN	
Spi-B	=>	Spi-B/NN	
Stat1	=>	Stat1/NN	

Stat3	=>	Stat3/NN	
SupT1	=>	SupT1/NN	◆cell line
T cell activation	=>	T/NN cell/NN activation/NN	
T cell	=>	T/NN cell/NN	
T cells	=>	T/NN cells/NNS	
T lymphocyte	=>	T/NN lymphocyte/NN	
T lymphocytes	=>	T/NN lymphocytes/NNS	
T lymphoma	=>	T/NN lymphoma/NN	
T-cell activation	=>	T-cell/NN activation/NN	
T-cell genes	=>	T-cell/NN genes/NNS	
T-cell	=>	T-cell/NN	
T-cells	=>	T-cells/NNS	
T3	=>	T3/NN	
TAL-1	=>	TAL-1/NN	
TATA box	=>	TATA/NN box/NN	
TCDD	=>	TCDD/NN	
TCF-1	=>	TCF-1/NN	
TCR2	=>	TCR2/NN	
TCR	=>	TCR/NN	
TGF-beta	=>	TGF-beta/NN	
THP-1	=>	THP-1/NN	◆cell line
THP1	=>	THP1/NN	◆cell line
TNF alpha	=>	TNF/NN alpha/NN	
TNF gene	=>	TNF/NN gene/NN	
TNF-alpha	=>	TNF-alpha/NN	
TNFR1	=>	TNFR1/NN	
TPA	=>	TPA/NN	
TRAF3	=>	TRAF3/NN	
TRP	=>	TRP/NN	◆cell line
Tax1	=>	Tax1/NN	
Tax	=>	Tax/NN	

Thymocyte	=>	Thymocyte/NN	
Thymocytes	=>	Thymocytes/NNS	
U937	=>	U937/NN	◆cell line
UT-7	=>	UT-7/NN	◆cell line
VCAM-1	=>	VCAM-1	
VD	=>	VD/NN	
VDR	=>	VDR/NN	
VitD3	=>	VitD3/NN	
WEHI-3B(D+)	=>	WEHI-3B(D+)/NN	◆cell line
Western blot	=>	Western/NN blot/NN	
Western blotting	=>	Western/NN blotting/NN	
X chromosome	=>	X/NN chromosome/NN	
X-chromosome	=>	X-chromosome/NN	
YT-N17	=>	YT-N17/NN	◆cell line
[3H]calcitriol/VDR	=>	[3H]calcitriol/VDR/NN	
[3H]calcitriol	=>	[3H]calcitriol/NN	
[3H]dexamethasone	=>	[3H]dexamethasone/NN	
[3H]thymidine	=>	[3H]thymidine/NN	
activated T cells	=>	activated/VBN T/NN cells/NNS	
acute	=>	acute/JJ	
adhesion molecules	=>	adhesion/NN molecules/NNS	
aldosterone	=>	aldosterone/NN	
alpha-methyl-1	=>	alpha-methyl-1/NN	
alpha,25-(OH)2D3	=>	alpha,25-(OH)2D3/NN	
alpha-tcp	=>	alpha-tcp/NN	
alpha-tosyl-L-lysine	=>	alpha-tosyl-L-lysine/NN	
alpha-tosylphenylalanyl	=>	alpha-tosylphenylalanyl/JJ	
androgen receptor gene	=>	androgen/NN receptor/NN gene/NN	
apoptosis	=>	apoptosis/NN	

atopic	=>	atopic/JJ
autosome	=>	autosome/NN
bcl-2	=>	bcl-2/NN
beta-globin	=>	beta-globin/NN
bp	=>	bp/NN
breast cancer	=>	breast/NN cancer/NN
bronchoalveolar	=>	bronchoalveolar/JJ
burst-forming unit-erythroid (BFU-E)	=>	burst-forming/JJ unit-erythroid/NN (/(BFU-E/NN)/)
c-Fos	=>	c-Fos/NN
c-Jun	=>	c-Jun/NN
c-Myc	=>	c-Myc/NN
c-Rel	=>	c-Rel/NN
c-fes	=>	c-fes/NN
c-fos gene	=>	c-fos/NN gene/NN
c-fos	=>	c-fos/NN
c-jun	=>	c-jun/NN
c-kit	=>	c-kit/NN
c-myb	=>	c-myb/NN
c-myc	=>	c-myc/NN
cAMP	=>	cAMP/NN
cDNAs	=>	cDNAs/NNS
calcineurin	=>	calcineurin/NN
calcium ionophore	=>	calcium/NN ionophore/NN
calcium	=>	calcium/NN
cell lines	=>	cell/NN lines/NNS
cells	=>	cells/NNS
cellular gene	=>	cellular/JJ gene/NN
chromatin	=>	chromatin/NN
corticosteroid	=>	corticosteroid/NN

cortisol	=>	cortisol/NN
cycloheximide	=>	cycloheximide/NN
cyclosporin A	=>	cyclosporin/NN A/NN
cysteine	=>	cysteine/NN
cytokine genes	=>	cytokine/NN genes/NNS
cytokine	=>	cytokine/NN
cytokines	=>	cytokines/NNS
cytoplasm	=>	cytoplasm/NN
dexamethasone	=>	dexamethasone/NN
differentiation	=>	differentiation/NN
downstream sequence	=>	downstream/JJ sequence/NN
egr-1	=>	egr-1/NN
electrophoretic mobility shift assays	=>	electrophoretic/JJ mobility/NN shift/NN assays/NNS
endothelial cells	=>	endothelial/JJ cells/NNS
enzyme	=>	enzyme/NN
expression	=>	expression/NN
factor-1	=>	factor-1/NN
fluoroquinolone	=>	fluoroquinolone/NN
forskolin	=>	forskolin/NN
gamma(c)	=>	gamma(c)
gelatinase	=>	gelatinase/NN
gene expression	=>	gene/NN expression/NN
gene transcription	=>	gene/NN transcription/NN
genes	=>	genes/NNS
globin genes	=>	globin/NN genes/NNS
globin	=>	globin/NN
glucocorticoid receptor	=>	glucocorticoid/NN receptor/NN
glucocorticoid receptors	=>	glucocorticoid/NN receptors/NNS
glucocorticoid	=>	glucocorticoid/NN

glucocorticoids	=>	glucocorticoids/NNS
granulocyte DNA	=>	granulocyte/NN DNA/NN
granulocytes	=>	granulocytes/NNS
hGR	=>	hGR/NN
hIL-5	=>	hIL-5/NN
hematopoietic cells	=>	hematopoietic/JJ cells/NNS
hemin	=>	hemin/NN
histone H5	=>	histone H5/NN
human A gamma-globin gene	=>	human/JJ A/NN gamma-globin/NN gene/NN
human T cells	=>	human/JJ T/NN cells/NNS
human immunodeficiency virus long terminal repeat	=>	human/JJ immunodeficiency/NN virus/NN long/JJ terminal/JJ repeat/NN
human immunodeficiency virus type 1 enhancer	=>	human/JJ immunodeficiency/NN virus/NN type/NN 1/CD enhancer/NN
human immunodeficiency virus type 1	=>	human/JJ immunodeficiency/NN virus/NN type/NN 1/CD
human immunodeficiency virus	=>	human/JJ immunodeficiency/NN virus/NN
human monocytes	=>	human/JJ monocytes/NNS
immediate early gene	=>	immediate/JJ early/JJ gene/NN
immediate-early gene	=>	immediate-early/JJ gene/NN
immune response	=>	immune/JJ response/NN
immune system	=>	immune/JJ system/NN
immunoglobulin genes	=>	immunoglobulin/NN genes/NNS
immunoglobulin	=>	immunoglobulin/NN
inflammatory genes	=>	inflammatory/JJ genes/NNS
inositol	=>	inositol/NN

1,4,5-trisphosphate/calcium		1,4,5-trisphosphate/calcium/N N
inositol phosphate	=>	inositol phosphate/NN
inositol-1,4,5-trisphosphate	=>	inositol-1,4,5-trisphosphate/N N
interferon	=>	interferon/NN
interleukin-2 enhancer	=>	interleukin-2/NN enhancer/NN
interleukin-2 gene	=>	interleukin-2/NN gene/NN
interleukin-2	=>	interleukin-2/NN
interleukin-4	=>	interleukin-4/NN
interleukin-5	=>	interleukin-5/NN
interleukin-6	=>	interleukin-6/NN
interleukin-8	=>	interleukin-8/NN
islet	=>	islet/NN
kappa B sequence	=>	kappa/NN B/NN sequence/NN
leukocyte	=>	leukocyte/NN
lining cell	=>	lining/JJ cell
lining region	=>	lining/JJ region
lipopolysaccharide	=>	lipopolysaccharide/NN
long terminal repeat	=>	long/JJ terminal/JJ repeat/NN
lymph-node	=>	lymph-node/NN
lymph	=>	lymph/NN
lymphocyte	=>	lymphocyte/NN
lymphocytes	=>	lymphocytes/NNS
lymphokine genes	=>	lymphokine/NN genes/NNS
lyn	=>	lyn/NN
mAb	=>	mAb/NN
mAbs	=>	mAbs/NNS
mRNA	=>	mRNA/NN
macrophages	=>	macrophages/NNS
mast cell	=>	mast/NN cell/NN

megakaryocyte	=>	megakaryocyte/NN
mg-1	=>	mg-1/NN
microM	=>	microM/NN
mineralocorticoid	=>	mineralocorticoid/NN
molecular	=>	molecular/JJ
molecule-1	=>	molecule-1/NN
monocytes	=>	monocytes/NNS
mononuclear leukocytes	=>	mononuclear/JJ leukocytes/NNS
myb	=>	myb/NN
necrosis	=>	necrosis/NN
nef	=>	nef/NN
neutrophils	=>	neutrophils/NNS
nuclear extracts	=>	nuclear/JJ extracts/NNS
nuclear factor kappa B	=>	nuclear/JJ factor/NN kappa/NN B/NN
nuclear factor of activated T cells	=>	nuclear/JJ factor/NN of/IN activated/VBN T/NN cells/NNS
nuclear factors	=>	nuclear/JJ factors/NNS
nuclear translocation	=>	nuclear/JJ translocation/NN
nucleotide	=>	nucleotide/NN
octamer motif	=>	octamer/NN motif/NN
octamer	=>	octamer/NN
okadaic acid	=>	okadaic/JJ acid/NN
oligoadenylate synthetase	=>	oligoadenylate/NN synthetase/NN
oncoprotein	=>	oncoprotein/NN
oriP	=>	oriP/NN
p100	=>	p100/NN
p105	=>	p105/NN
p38	=>	p38/NN

p50	=>	p50/NN	
p52	=>	p52/NN	
p65	=>	p65/NN	
patient	=>	patient/NN	
patients	=>	patients/NNS	
peptide	=>	peptide/NN	
peri-kappa B	=>	peri-kappa/NN B/NN	
peripheral blood	=>	peripheral/JJ blood/NN	
peroxisome	=>	peroxisome/NN	◆sub-cell source
phagocyte	=>	phagocyte/NN	
phorbol ester	=>	phorbol/NN ester/NN	
phorbol esters	=>	phorbol/NN esters/NNS	
phosphate	=>	phosphate/NN	
phosphatidylinositol	=>	phosphatidylinositol/NN	
phytohemagglutinin	=>	phytohemagglutinin/NN	
plaque	=>	plaque/NN	
plasma cortisol	=>	plasma/NN cortisol/NN	
platelet-activating factor	=>	platelet-activating/JJ factor/NN	
pp56(lck)	=>	pp56(lck)/NN	
poly(ADP-ribose)	=>	poly(ADP-ribose)/JJ	
polymerase	=>	polymerase/NN	
prednisolone	=>	prednisolone/NN	
prodigiosin	=>	prodigiosin/NN	
proliferation	=>	proliferation/NN	
promoter	=>	promoter/NN	
promoter	=>	promoter/NN	
promoters	=>	promoters/NNS	
promoters	=>	promoters/NNS	
protein kinase C	=>	protein/NN kinase/NN C/NN	

protein-protein	=>	protein-protein/JJ
proto-oncogenes	=>	proto-oncogenes/NNS
protooncogene c-fos	=>	protooncogene/NN c-fos/NN
rapamycin	=>	rapamycin/NN
receptors	=>	receptors/NNS
sIgD	=>	sIgD/NN
sIgM	=>	sIgM/NN
signaling pathway	=>	signaling/NN pathway/NN
signaling pathways	=>	signaling/NN pathways/NNS
signaling receptor	=>	signaling/NN receptor
signalling pathway	=>	signalling/NN pathway/NN
signalling pathways	=>	signalling/NN pathways/NNS
sn-1,2 dioctanoyl glycerol	=>	sn-1,2/NN dioctanoyl/JJ glycerol/NN
steroid	=>	steroid/NN
substrates	=>	substrates/NNS
tal-1	=>	tal-1/NN
therapy	=>	therapy/NN
thrombin	=>	thrombin/NN
thymocyte	=>	thymocyte/NN
thymocytes	=>	thymocytes/NNS
transcription factor	=>	transcription/NN factor/NN
transcription factors	=>	transcription/NN factors/NNS
transcription	=>	transcription/NN
transcriptional activation	=>	transcriptional/JJ activation/NN
transgenic mice	=>	transgenic/JJ mice/NNS
tumor necrosis factor alpha	=>	tumor/NN necrosis/NN factor/NN alpha/NN
tumor	=>	tumor/NN
tumour	=>	tumour/NN

tyrosine kinase	=>	tyrosine/NN kinase/NN
tyrosine phosphorylation	=>	tyrosine/NN phosphorylation/NN
tyrosine	=>	tyrosine/NN
upstream sequence	=>	upstream/JJ sequence/NN
v-abl	=>	v-abl/NN
v-abl	=>	v-abl/NN
vitamin	=>	vitamin/NN

MEDLINE Database Flags

UI - => UI/LS -/:

TI - => TI/LS -/:

AB - => AB/LS -/:

Appendix 2

-ed

The column "Prenominal" shows the POS that should be assigned to the word when it is a prenominal modifier. The column "After "be" without "by"- " shows the POS that should be assigned to the word when it is a main verb of passive sentence without the by-phrase that serves as the semantic subject.

	Prenominal	After "be" without "by-"
(co)cultured	VBN	VBN
(co)localized	JJ	JJ
abolished	VBN	VBN
accepted	VBN	VBN
accumulated	VBN	VBN
acetylated	VBN	VBN
achieved	VBN	VBN
acquired	VBN	VBN
activated	VBN	VBN
added	VBN	VBN
administered	VBN	VBN
admitted	VBN	VBN
advanced	JJ	??
affected	VBN	VBN
aged	JJ	JJ
altered	JJ?	VBN?
amplified	VBN	VBN
analysed	VBN	VBN
analyzed	VBN	VBN

anchored	VBN	VBN
energized	VBN	VBN
anucleated	JJ	JJ
applied	VBN	VBN
arranged	VBN	VBN
arrested	VBN	VBN
ascribed	VBN	VBN
assessed	VBN	VBN
assigned	VBN	VBN
associated	VBN	VBN
assumed	VBN	VBN
attenuated	VBN	VBN
augmented	JJ	JJ
balanced	JJ	JJ
barbed	JJ	JJ
believed	VBN	VBN
biotinylated	VBN	VBN
blocked	VBN	VBN
blunted	VBN	VBN
born	JJ	JJ
bound	VBN	VBN
butylated	VBN	VBN
called	VBN	VBN
capped	JJ	??
changed	VBN	VBN
characterized	VBN	VBN
charged	JJ	JJ
classified	VBN	VBN
cloned	VBN	VBN

co-transfected	VBN	VBN
coexpressed	VBN	VBN
coimmunoabsorbed	VBN	VBN
coinfectd	VBN	VBN
combined	JJ	VBN
committed	VBN	VBN
compared	VBN	VBN
compensated	VBN	VBN
composed	VBN	VBN
comprised	VBN	VBN
computerized	VBN	VBN
concluded	VBN	VBN
conditioned	JJ	JJ
conferred	VBN	VBN
confined	VBN	VBN
confirmed	VBN	VBN
conserved	VBN	VBN
considered	VBN	VBN
consisted	VBN	VBN
constructed	VBN	VBN
contained	VBN	VBN
continued	JJ	VBN
controlled	JJ	VBN
converted	VBN	VBN
coordinated	VBN	VBN
correlated	VBN	VBN
coselected	VBN	VBN
cotransfected	VBN	VBN
coupled	JJ	VBN

created	VBN	VBN
crosslinked	VBN	VBN
cultured	VBN	VBN
debulked	VBN	VBN
decreased	VBN	VBN
defined	VBN	VBN
degraded	VBN	VBN
delayed	VBN	VBN
deleted	JJ	VBN
delimited	VBN	VBN
demonstrated	VBN	VBN
dephosphorylated	VBN	VBN
derived	VBN	VBN
described	VBN	VBN
deseased	JJ	JJ
designated	VBN	VBN
designed	VBN	VBN
detailed	JJ	VBN
detected	VBN	VBN
determined	VBN	VBN
developed	VBN	VBN
differentiated	VBN	VBN
diminished	VBN	VBN
directed	VBN	VBN
discussed	VBN	VBN
disordered	VBN	VBN
displayed	VBN	VBN
disrupted	VBN	VBN
dissociated	VBN	VBN

diverged	JJ	JJ	
documented	VBN	VBN	
downregulated	??	VBN	
duplicated	JJ	JJ	
effected	VBN	VBN	
elaborated	VBN	VBN	
elevated	JJ	JJ	
eliminated	VBN	VBN	
elucidated	VBN	VBN	
elutriated	VBN	VBN	
embedded	JJ	JJ	
employed	VBN	VBN	
encoded	VBN	VBN	
endocytosed	VBN	VBN	chemical reaction
endowed	JJ	JJ	
engineered	VBN	VBN	
enhanced	VBN	VBN	
enhanced	VBN	VBN	
enriched	JJ	JJ	
erected	JJ	?	
established	JJ	VBN	
esterified	VBN	VBN	
evaluated	VBN	VBN	
examined	VBN	VBN	
excised	VBN	VBN	
excluded	VBN	VBN	
exerted	VBN	VBN	
expanded	VBN	VBN	
exposed	VBN	VBN	

expressed	VBN	VBN	
extended	VBN	VBN	
extracted	VBN	VBN	
grown	JJ	VBN	
localized	JJ	JJ	
	Prenominal	After "be" without "by-"	
fixed	JJ	?	
footprinted	VBN	VBN	
forced	VBN	VBN	
formed	VBN	VBN	
formylated	VBN	VBN	
formylated	VBN	VBN	
found	VBN	VBN	
fragmented	JJ	JJ	
generalized	VBN	VBN	
generated	VBN	VBN	
given	JJ	VBN	Given/IN XXX (conditional)
grown	JJ	VBN	
hypopigmented	JJ	JJ	
hypothesized	VBN	VBN	
identified	VBN	VBN	
immobilized	VBN	VBN	
immortalized	VBN	VBN	
immunostained	VBN	VBN	
impaired	JJ	JJ	
implicated	VBN	VBN	
increased	VBN	VBN	
incubated	VBN	VBN	

induced	VBN	VBN	
infected	JJ	VBN	
infected	JJ	VBN	
inflamed	JJ	JJ	when it modifies tissues or organs
inhaled	VBN	JJ	
inherited	VBN	VBN	
inhibited	VBN	VBN	
initiated	VBN	VBN	
injected	VBN	VBN	
inoculated	VBN	VBN	
integrated	VBN	VBN	
interconnected	JJ	JJ	
inverted	JJ VBN?	VBN	
investigated	VBN	VBN	
involved	VBN	VBN	
isolated	VBN	VBN	
juxtaposed	VBN	VBN	
	Prenominal	After "be" without "by-"	
known	JJ	VBN	
labelled	VBN	VBN	
limited	JJ	JJ	
linked	VBN	VBN	
loaded	VBN	VBN	
localized	JJ	JJ	
located	JJ	JJ	
lowered	JJ	JJ	
maintained	JJ	JJ	

mapped	VBN	VBN	
marked	JJ	JJ	
measured	VBN	VBN	
mediated	VBN	VBN	
microdissected	VBN	VBN	
mixed	JJ	VBN	
modified	VBN	VBN	
modulated	VBN	VBN	
mutagenized	VBN	VBN	
mutated	VBN	VBN	
myristoylated	VBN	VBN	
named	VBN	VBN	
needed	VBN	VBN	
neutralized	VBN	VBN	
noted	VBN	VBN	
noticed	VBN	VBN	
nucleated	JJ	JJ	
observed	VBN	VBN	
obtained	VBN	VBN	
opsonized	VBN	VBN	
optimized	VBN	VBN	
overexpressed	VBN	VBN	
oxidized	VBN	VBN	
	Prenominal	After "be" without "by-"	
paired	JJ	JJ	
paralleled	JJ	JJ	
parasitized	VBN	VBN	
performed	VBN	VBN	

perturbed	VBN	VBN	
phosphorylated	VBN	VBN	
plated	VBN	VBN	
poised	VBN	VBN	
polarized	VBN	VBN	
polyadenylated	VBN	VBN	
polyubiquitinated	VBN	VBN	
potentiated	JJ	JJ	
predicted	VBN	VBN	
prepared	VBN	VBN	
presented	VBN	VBN	
preserved	JJ	JJ	
presumed	VBN	VBN	
primed	VBN	VBN	
processed	VBN	VBN	
produced	VBN	VBN	
programmed	VBN	VBN	
prolonged	JJ	JJ	
pronounced	JJ	JJ	
proposed	JJ	VBN?	
protected	JJ	VBN	
published	VBN	VBN	
pulsed	JJ	JJ?	
purified	VBN	VBN	
radiolabeled	VBN	VBN	
reactivated	VBN	VBN	
reared	VBN	VBN	
rearranged	VBN	VBN	
reasoned	VBN	VBN	

recognized	VBN	VBN	
reconstituted	VBN	VBN	
reconstructed	VBN	VBN	
recovered	VBN	VBN	
recruited	VBN	VBN	
reduced	VBN	VBN	
referred	VBN	VBN	
regulated	VBN	VBN	
related	JJ	JJ	
released	VBN	VBN	
rendered	VBN	VBN	
repeated	JJ	VBN	
replaced	VBN	VBN	
replicated	VBN	VBN	
reported	VBN	VBN	
required	JJ	VBN	
rescued	VBN	VBN	
resected	JJ	VBN	
restricted	JJ	JJ	
resulted	VBN	VBN	
retained	VBN	VBN	
retarded	VBN	VBN	
retrodiffentiated	VBN	VBN	
returned	VBN	VBN	
revealed	VBN	VBN	
reversed	VBN	VBN	
revisited	VBN	VBN	
ruptured	VBN	VBN	
saturated	JJ	JJ	

secreted	VBN	VBN	
seeded	VBN	VBN	
selected	VBN	VBN	
sensitized	VBN	VBN	
sequestered	VBN	VBN	
seroconverted	VBN	VBN	
shared	JJ	VBN	
skewed	JJ	VBN	
slipped	JJ	JJ	
sorted	VBN	VBN	
speckled	JJ	JJ	
spliced	VBN	VBN	
stimulated	VBN	VBN	
stranded	JJ	JJ	
stratified	depending on the context		
stressed	JJ	JJ	
studied	VBN	VBN	
submitted	VBN	VBN	
substituted	VBN	VBN	
subtracted	VBN	VBN	
suggested	VBN	VBN	
supercoiled	JJ	JJ	
suppressed	VBN	VBN	
surveyed	VBN	VBN	
sustained	JJ	JJ	
synchronized	VBN	VBN	
synthesized	VBN	VBN	
targeted	VBN	VBN	
targetted	VBN	VBN	

termed	VBN	VBN	
tested	VBN	VBN	
transactivated	VBN	VBN	
transcribed	VBN	VBN	
transduced	VBN	VBN	
transfected	VBN	VBN	
transferred	VBN	VBN	
transformed	VBN	VBN	
translated	VBN	VBN	
translocated	VBN	VBN	
transplanted	VBN	VBN	
treated	VBN	VBN	
truncated	VBN	VBN	
	Prenominal	After "be" without "by-"	
ubiquitinated	VBN	VBN	
ubiquitinylated	VBN	VBN	
underlined	VBN	VBN	
underphosphorylated	VBN	VBN	
undertaken	VBN	VBN	
used	VBN	VBN	
utilized	VBN	VBN	
washed	VBN	VBN	
XXXmerized	JJ	JJ	dimerized, polymerized, etc

Appendix 3

-ing

JJ (those marked with * has an adjective entry/entries in LDOCE)

- adhering/JJ cells
- adhering/JJ cells
- circulating/JJ levels of cortisol
- circulating/JJ levels of cortisol
- clipping/JJ bandshift assays
- corresponding/JJ NN (all occurrences in WSJ were tagged as JJ)
- cross-linking/JJ CD40
- cutting/JJ edge
- cytoreductive staging/JJ laparotomy
- determining/JJ role
- housekeeping/JJ genes
- impeding/JJ NN
- infiltrating/JJ ductal carcinomas (when "infiltrating" modifies cancer-related nouns to describe the nature of the cancer)
- it is tempting/JJ* to ...
- opposing/JJ effects (as this does not mean an effect called "opposing")
- preexisting/JJ*, latent, STAT1-SATAT2 signaling complex
- presenting/JJ HLA molecules
- priming/JJ cultures (meaning "primary")
- proliferating/JJ state
- cytoreductive staging/JJ laparotomy
- striking/JJ {similarities, differences etc}/NN
- surrounding/JJ* NN (only before noun)
- it is tempting/JJ* to ...
- underlying/JJ* difference
- unifying/JJ hypothesis
- working/JJ* hypothesis

NN

- non-DNA/JJ binding/NN sequence (to be consistent with “DNA/NN binding/NN sequence”)
- blocking/NN factor
- clustering/NN algorithm (method)
- comparative mapping/NN data
- docking/NN sites
- reading/NN frame
- blocking/NN factor
- signaling/NN cascade
- nuclear targeting/NN sequence
- testing/NN data

VBG

- activating/VBG Gi proteins
- activating/VBG region
- multiple activating/VBG agents
- alternating/VBG RCGCRYGCGY consensus
- binding/VBG protein (motif, site)
- blocking/VBG Abs
- budding/VBG yeast
- circulating/VBG human lymphocytes
- coding/VBG region
- crossreacting/VBG? proteins
- deactivating/VBG cytokine
- developing/VBG NN
- differentiating/VBG agent
- differentiating/VBG hematopoietic progenitors
- following/VBG NN (when the entire phrase serves as an adverbial. If it is not, “the following/JJ NN”)
- homodimerizing/VBG cytokine receptors
- internalizing/VBG antibodies
- intervening/VBG NN
- intracellular signaling/VBG agents
- ionizing/VBG radiation
- limiting/VBG element
- rate limiting/VBG

- comparative mapping/NN data
- mediating/VBG receptor–G protein
- menstruating/VBG women (patients)
- multiple activating/VBG agents
- overlapping/VBG endothelium
- proliferating/VBG cells
- resting/VBG monocytes (cells, etc)
- signaling/VBG molecules (agents, proteins, etc)
- smoking/VBG patients
- splicing/VBG NN
- stimulating/VBG T cell blasts
- transforming/VBG growth factor
- ubiquitinating/VBG enzymes