Given a base word form, the task is to assign the appropriate WordNet sense to every instance of that form appearing in the WordNet glosses. The Mantag interface will present you with a new base form from a centralized list, and that form may be a single word or a collocation. A collocation is a sequence of two or more words expressing a single concept which, as a compound, is defined in WordNet. These are things like compound nouns, idioms, or phrasal verbs. For example, **snow eater** is a noun collocation in WordNet (it's a dry, warm wind), as is **flower child**, **kick the bucket** is an idiom, and **eat out** is a phrasal verb.

Note that you are no longer tagging all derived forms of the base form, and are no longer required to tag all collocations containing that base form!

For each base form, the Mantag interface display all glosses that *may* contain the base form. You will see one gloss at at time. A gloss is a definition, and the word or words it defines is presented on the line labeled Synset, right above the gloss. The idea is that the context given by the gloss along with the relations for the Synset should aid you in disambiguating the term you are tagging.

1. Read the entire gloss, checking for whether the highlighted word/collocation is the one you are searching for, or an inflected form of it. Inflected forms are the different tenses of verbs and the plural forms of nouns ("ate" is an inflected form of **eat**, "children" is an inflected form of **child**). Take the word **cut**. The first "cuts" in the examples below and the first 2 instances of "cutting" are inflected forms of the verb. The second "cuts" is an inflected form of the noun **cut**. "cutter" is a noun, derived from the verb **cut.** The last instance of "cutting", however, is not an inflected form of the verb "to cut", but rather a noun derived from it.

    i.   a machine that <cuts> grain and binds it in sheaves
    ii.   an implement for <cutting> bolts
    iii.   he was punished for taking too many <cuts> in his math class
    iv.   a <cutter> (tool for <cutting>) for floor tiles
    v.   a <cutting> of a plant

If you are searching for a collocation, check now for whether the gloss does indeed contain an instance of that collocation, and is not a false hit. If it is not a false hit, determine if the collocation has already been "globbed". Glob it if not. Possible parts of collocations will appear with a light green background, with the first word highlighted in yellow. They may or may not form a collocation with your search term. For example, **cut glass** is a noun collocation for a kind of decorated glass. The word "glass" will appear with a green background if the gloss is something like any of the examples below:

    vi.   a vase made of cut glass
    vii.   to cut your finger on a piece of glass
    viii.   a procedure in which you cut glass

Only the first example is an instance of the noun collocation **cut glass**, the others are false hits.

The parts of a collocation may be separated by one or more words. This is particularly true of verb collocations, like **cut away**.

    ix.   cut away the branch
    x.   cut the branch away
    xi   cut the branches, hanging vines, and icky spider webs away

All of the above are occurrences of the verb collocation **cut away**.

Also be on the lookout for whether the word or collocation was accidentally globbed as part of a collocation that it does not belong to. Say you are tagging the noun collocation **black bear**. The following gloss contains an instance of this collocation, but the "bear" part has been globbed previously (and erroneously) to the verb collocation **bear in mind**.

    xii.   while in the woods, it's good to keep the black bear in mind

In this case, "bear in mind" should be unglobbed, so that globbing and tagging of **black bear** can proceed.

    ***   Note that not all occurrences of a collocation should be globbed! For   ***
    ***   example, "black bear" could be a bear who happens to be black OR it   ***
    ***   could be a kind of bear (as opposed to a grizzly, or polar bear). The   ***
    ***   context of the sentence should tell you which it is. For example,   ***

    xiii.   Was it a black bear or a brown one that you spotted in the woods? [not a glob]
    xiv.   reddish-brown color phase of the American black bear [is a glob]

For noun collocations, it is usually the case that the parts of the collocation cannot be split up. This is a useful indication that you have a glob. For example, in the first sentence, you can reword the sentence so that "black" and "bear" are separated: "the bear I spotted in the woods was black". In the second sentence you cannot: "the American bear was black".

2. Check the gloss for other cases of bad globs which aren't necessarily related to the word you are working on. If you see something wrongly globbed, then unglob it!

3. Determine part of speech for the word/collocation. This is not always as obvious as it seems! There are four particularly tricky cases. These are all tricky because the part of speech of a word is not always the same as the grammatical function that the word is performing in the sentence or phrase. For instance, nouns can function similarly to what are traditionally called adjectives, and verbs can take on the roles of nouns or adjectives.

a) Adjective vs. noun modifying a noun

The usual modifier for a noun is an adjective preceding it:

xv.   <favorable> conditions are forecast
xvi.   <damp> weather is behind us

And sometimes after the noun:

xvii.   conditions <unfavorable> for flying

Nouns can also serve as modifiers, similar to adjectives:

xviii. <entertainment> value
xix.   <children's> books
xx.   <cotton> shirts

The general rule of thumb for deciding whether it is a noun or adjective is to check the sense list first for whether there is an adjective sense corresponding to the word, and, if not, then whether there's a corresponding noun sense. So, "damp" in "damp weather" is an adjective, even though a noun sense exists. And "cotton" in "cotton shirts" is a noun, which is modifying another noun.

If there is no adjective sense in WordNet, then you should make sure that it is not truly an adjective that is missing from WordNet. A good clue that you have an adjective is if you try to modify it with "very" or "rather" and it sounds ok: "very/rather favorable conditions" (ok) vs. "very/rather cotton shirts" (not ok). Another good clue is if you can make a comparative or superlative form out of it ("damper/dampest/more favorable/most favorable conditions" are all adjectives, but "cottoner/cottonest/more cotton/most cotton shirts" are not).

If either of these tests come up ok (that is, "very/rather x" sounds good or either "x-er/x-est" or "more/most x" sounds good), and there is no matching adjective sense, then assign "sense not in WordNet".

    ***   Note that these tests are only valid if they come up ok. Then you   ***
    ***   know you have an adjective for sure. If the tests are *not* ok, then it   ***
    ***   may still be an adjective. This is because the tests only work for   ***
    ***   certain kinds of adjectives, but not all.   ***

If the tests are not ok (that is, none of "very/rather x" and "x-er/x-est/more x/most x" sound good), then check for a matching noun sense. If there is no matching noun sense, then do not assign *any* sense. (But see b and c below first regarding present and past participles, since it might be a verb!). If a noun sense does exist, then the word can be considered a noun, and be tagged to the noun sense.

The noun-sense rule applies only when the word is modifying a noun. If the word is being used predicatively (that is, after some form of the verb "to be", or where the verb could be

replaced by a verb such as seem, look, appear, etc.) In the predicative case, there may be some confusion as to whether what follows the verb is an adjective or a noun. So, in

      xxi.   the weather was <damp>

"damp" is an adjective here. Notice that you can replace "was" with seemed/looked/appeared and still get a grammatical sentence: "the weather seemed/looked/appeared damp". But note the difference between the pairs:

      xxii.   he was <drunk> [function = adjective]
      xxiii.   he was really <drunk> [function = adjective]

and

      xxiv.   he was a <drunk> [function = noun]
      xxv.   he was obviously a <drunk> [function = noun]

In the second pair, **drunk** is clearly a noun, not an adjective. It is the complement of the verb "to be" here. Two reliable ways to recognize a noun are if it is (or can be) preceded by a determiner (such as "a" or "the") or adjective (he was a **silly** drunk).

In sum:
- When you have the situation of "modifier noun", the modifier will be an *adjective* when
  - there is a corresponding adjective sense in WordNet for the meaning it is being used with <u>or</u>
  - there is no corresponding adjective sense, but any of the the very/rather (sounds good when you preface it with "very" or "rather") or –er/– est/more/most (x-er/x-est/more x/most x) tests come up ok, in which case "Sense not in WordNet" should be assigned
- The modifier will be a noun when
  - There is no corresponding adjective, <u>and</u> the very/rather and –er/–est/more x/most x tests are not ok, <u>and</u> a corresponding noun sense exists
- If it is neither an adjective nor a noun, it might be a verb (see b and c below)

<u>b) Adjective vs. present participle (-ing form) of verb</u>

The -ing form of verbs can function as adjectives. For instance,

      xxvi.   you are <frightening> me [function = verb]
      xxvii.   that is a <frightening> prospect [function = adjective/modifier]
      xxviii. that prospect is <frightening> to me [function = adjective/modifier]
      xxix.   that prospect is <frightening> me [function = verb]

xxx.   we are <working> hard today [function = verb]
xxxi.   the car is in good <working> order [function = adjective/modifier]

How to tell? The easy case is when the word is modifying a noun. In general, these are adjectives if there is a corresponding adjective sense in WordNet. Such adjective senses exist for **frightening** and **working**. However, this is not the case for "clicking" and "playing", so that in the following sentences,

xxxii. the door opened with a <clicking> sound
xxxiii. the baseball player's <playing> record was impeccable

the appropriate verb senses of **click** and **play** would be selected instead. (This is because these are verbs playing the part of adjectives, but are not adjectives in themselves.)

When the word appears predicatively (after some form of the verb "to be"), the rule can't always be applied since it might be impossible to tell whether it is being used as a verb or an adjective.

xxxiv. That dress is <striking> on you [adjective]
xxxv.   The workers are <striking> for better working conditions [verb]
xxxvi. The women are <striking> [ambiguous]

Without more information, you cannot know whether the third sentence means that the women are picketing, or whether they are beautiful. For ambiguous cases like this, do not assign a sense, and the lexicographers will make the determination.

c) Adjective vs. past participle (usually -en form) of verb

Past tense participles can also function as adjectives. The past tense participle is the form of the verb that appears with the past tense auxiliary "have". It usually, though not always, ends in -en or -ed: "written", "destroyed", and "spun" are past participles of write, destroy and spin, respectively.

The rule of thumb will be similar to the present participle cases. Where the word modifies a noun, check first for a corresponding adjective sense. If no adjective sense exists, then assign the verb sense (if there is one that matches the meaning as used in the sentence).

xxxvii. that wasn't the <intended> result [adjective]
xxxviii.make out an <itemized> list [verb functioning as adjective]

Again, the hard cases occur when the word appears predicatively (ie., after some form of the verb "to be", or where the preceding verb can be replaced by a verb such as seem/look/appear, etc.).

xxxix. The sentence was <written> down for clarity. [verb]

xl.   The sentence was <written> as opposed to spoken. [adjective]
xli.   The sentence was <written> [ambiguous]

In the first sentence, "written" is a verb. A good test of this is to put the auxiliary verb in the progressive – "The sentence WAS BEING written down for clarity". That makes it clear it is an act or action that occurred. The second sentence cannot be phrased that way and still have the same sense: "The sentence was being written as opposed to spoken".

In the third sentence, it is not clear whether "written" refers to an act of writing, or the attribute or quality of being written. For ambiguous cases like this, do not assign a sense, and the lexicographers will make the determination.

d) Noun vs. present participle (-ing form) of verb

To complicate things further, the present participle of verbs can function as a noun. Often, the distinction is easy to make, if it appears where a noun is called for grammatically, and there is a corresponding noun sense in WordNet.

xlii.   he made a <killing> in the stock market [noun]
xliii.   the discordant <ringing> of nonmusical metallic objects [noun]

If no noun sense exists, then assign the verb sense, if one exists, as for:

xliv.   The merry <frolicking> of the lark [verb functioning as noun]

However, if the word is being used as a verb, then a noun sense should never be assigned! This is easy if there is no noun sense, as for **frolicking**:

xlv.   the lark is given to <frolicking> merrily [verb]

or when it is obviously depicting an ongoing action:

xlvi.   make a sound like a car engine that is <firing> too soon [verb]

You can test this out, too. A verb can never be modified by "a" or "the" or a possessive pronoun such as my/your/our, etc. Try it with the 2 sentences above--it hurts!

But, again, there will be cases where this determination will be impossible to make:

xlvii.   She dislikes <writing> thank-you letters [verb]
xlviii.   She had no knowledge of the <writing> of the letter [noun]
xlix.   She had been talking about <writing> [ambiguous]

It is not clear whether "writing" in the 3rd sentence refers to the act of writing something (eg, a letter), or whether writing is the object itself (ie, her writing, or an author's writing,

marks on a piece of paper, etc.) For ambiguous cases like this, do not assign a sense, and the lexicographers will make the determination.

4. Determine sense (or senses).

In WordNet, senses are in part defined by their relations to other senses. For this reason, the WordNet relations can be very useful in narrowing down which of the senses applies to a particular occurrence of the form. The relations for any word or collocation can be viewed through the WordNet browser. Open another window and call up the browser by typing wnb <enter>. From the WordNet entry of the word you are tagging, clicking on one of the part of speech buttons will display a menu listing the kinds of relations that you can look at. The main ones that will be of help are Hypernyms, Derivationally related forms, and Domain. Not all relations will appear for all words and all parts of speech for a word. If you select one of the relations for a part of speech, the window will display only those senses that have that relation. You can print this out [via "Print current display" from the File menu].

Hypernym (ISA relation)

The immediate hypernym is the most relevant one here. It is the first indented relation just below the definition (preceded by an arrow =>). The hypernym relation will tell you what kind of thing (object or action) the word refers to. The higher up you go in the hypernym relations, the more general the senses get (and so often less informative). There is a new indentation for each level up you go. For instance, two senses of the noun **center** that are rather close are Senses 1 and 4:

Sense 1
center, centre, middle, heart, eye -- (an area that is approximately central within some larger region; "it is in the center of town"; "they ran forward into the heart of the struggle"; "they were in the eye of the storm")

and

Sense 4
center, centre, midpoint -- (a point equidistant from the ends of a line or the extremities of a figure)

If you look at the hypernyms for the noun senses of "center", you can see that Sense 1 is an area while Sense 4 is a point, what they have in common is a notion of centrality. Both are at some level locations, and eventually all nouns are entities (so that knowing that something is a kind of entity is not of much help at all!).

Sense 1
center, centre, middle, heart, eye -- (an area that is approximately central within some larger region; "it is in the center of town"; "they ran forward into the heart of the struggle"; "they were in the eye of the storm")

  => area, country -- (a particular geographical region of indefinite boundary (usually serving some special purpose or distinguished by its people or culture or geography); "it was a mountainous area"; "Bible country")
   => region -- (a large indefinite location on the surface of the Earth; "penguins inhabit the polar regions")
    => location -- (a point or extent in space)
     => object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
      => physical entity -- (an entity that has physical existence)
       => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 4
center, centre, midpoint -- (a point equidistant from the ends of a line or the extremities of a figure)
  => point -- (the precise location of something; a spatially limited location; "she walked to a point where she could survey the whole street")
   => location -- (a point or extent in space)
    => object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
     => physical entity -- (an entity that has physical existence)
      => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Derivationally related forms (related words in different parts of speech)

The verb **cut** has 41 senses, quite a lot to slog through each time you tag a verb. Good use of the derivational links (where they apply) can give you a shortcut through all those senses and do the selecting for you. Pretty nice, but how?

If you look at the derivational links (the RELATED TO-> pointer) for the verb **cut**, you will see that 7 of the 41 senses are related to senses of the nouns **cutter** and **cutting**. In particular, notice that only one sense is related to the noun **cutter** (it's related to 2 senses of **cutter**, the sense of the person who cuts, and the tool that does the cutting).

Sense 1
cut -- (separate with or as if with an instrument; "Cut the rope")
  RELATED TO->(noun) cutter#3
   => cutter -- (someone whose work is cutting (as e.g. cutting cloth for garments))
  RELATED TO->(noun) cutter#6
   => cutter, cutlery, cutting tool -- (a cutting implement; a tool for cutting)
  RELATED TO->(noun) cutting#3
   => cut, cutting -- (the act of cutting something into parts; "his cuts were skillful"; "his cutting of the cake made a terrible mess")

At some point in the tagging, you will get to the gloss for each of these nouns (since they each have some form of **cut** in their definitions). Since you know that sense #1 of the verb **cut** is related to sense #3 of the noun **cutter** (the first relation under **cut**), then when

you are tagging the gloss, you can simply assign sense #1 of the verb **cut** to both occurrences of the verb "cutting", without having to look further.

Domain (is this term restricted to one topic or area or field or context?)

Where they exist, the domain relations can be quite helpful in narrowing senses down. A word's domain will tell you whether it is restricted to some field or area such as Law or Art. Take the noun **work**. It has 7 senses, and if you look at its domains, you can see that one of its senses is restricted to the domain of physics, having to do with the transfer of energy (domain is depicted with the CATEGORY-> pointer).

Sense 5
work -- ((physics) a manifestation of energy; the transfer of energy from one physical system to another expressed as the product of a force and the distance through which it moves a body in the direction of that force; "work equals force times distance")
    CATEGORY->(noun) physics#1, physical science#1, natural philosophy#1

When tagging the form **work**, at some point the gloss for **energy** will appear, since it contains the word "work":

energy -- (the capacity of a physical system to do **work**; the units of energy are joules or ergs;)

This sense of **energy** is also in the domain of physics. That can be seen by the Domain line just below the Find box in the Mantag interface. The domain for the gloss you are tagging appears on that line. So it is very likely (and in this case, quite true) that the sense of **work** in the gloss above is the sense from physics.