I made a data dir following the recommendation of the book “Python Text Processing with NLTK 2.0 Cookbook” <https://www.packtpub.com/article/python-text-processing-nltk-20-creating-custom-corpora> as shown below:

NLTK defines a list of data directories, or paths, in *nltk.data.path*. Our custom corpora must be within one of these paths so it can be found by NLTK. So as not to conflict with the official data package, we'll create a custom *nltk\_data* directory in our home directory. Here's some Python code to create this directory and verify that it is in the list of known paths specified by *nltk.data.path*:

>>> import os, os.path  
>>> path = os.path.expanduser('~/nltk\_data')  
>>> if not os.path.exists(path):  
... os.mkdir(path)  
>>> os.path.exists(path)  
True  
>>> import nltk.data  
>>> path in nltk.data.path  
True

Then C:\Documents and Settings\frey\Application Data\nltk\_data\corpora\senseval (on Windows) is created, the corpora dir will have about 50 listed.

Delete everything from the senseval dir (about 5 .pos files) and copy the new senseval-3 files into that same dir. I named the senseval-3 files .pos, even though they don’t have pos tags (senseval-2 had pos and word sense <head> tags). That’s because the senseval import looks for .pos files for each word.

Then you can import the files like so:

>>> from nltk.corpus import senseval

>>> senseval.fileids()

['sense\_activate.v.pos', 'sense\_add.v.pos', 'sense\_appear.v.pos', 'sense\_argument.n.pos', 'sense\_arm.n.pos', 'sense\_ask.v.pos', 'sense\_atmosphere.n.pos', 'sense\_audience.n.pos']

>>> senseval.instances('sense\_activate.v.pos')

[SensevalInstance(word='activate.v', position=55, context=['Do', 'you', 'know', 'what', 'it', 'is', ',', 'and', 'where', 'I', 'can', 'get', 'one', '?', 'We', 'suspect', 'you', 'had', 'seen', 'the', 'Terrex', 'Autospade', ',', 'which', 'is', …

One thing to note if we write a new import, some training examples list a word with multiple senses, unlike our previous pos tagging, where one word would have one tag. An example like so:

<instance id="ask.v.bnc.00000105" docsrc="BNC">

<answer instance="ask.v.bnc.00000105" senseid="238102"/>

<context>

What if I find certain issues or situations difficult ? Your course leader will be available to help you . You are also <head>asked</head> to keep your church leaders informed of your involvement so that they can ensure you are adequately supported . After every client visit you are asked to call the office so that you can report how the visit went . We also hold regular meetings of volunteers to discuss issues of concern and encourage one another .

</context>

</instance>

<instance id="ask.v.bnc.00000106" docsrc="BNC">

<answer instance="ask.v.bnc.00000106" senseid="238102"/>

<answer instance="ask.v.bnc.00000106" senseid="238103"/>

<context>

Your course leader will be available to help you . You are also asked to keep your church leaders informed of your involvement so that they can ensure you are adequately supported . After every client visit you are <head>asked</head> to call the office so that you can report how the visit went . We also hold regular meetings of volunteers to discuss issues of concern and encourage one another . Volunteers visiting an ACET client have immediate access to professional nursing support through our 24 hour on call facility .

</context>

</instance>

Note that the first one has one sense (238102), while the second one has two senses (with just a single word tagged).