The Groningen Meaning Bank (GMB) project develops a corpus with rich syntactic and semantic annotations.
Annotations in GMB are generated semi-automatically and stem from two sources:
(i) Initial annotations from a set of standard NLP tools,
(ii) Corrections/refinements by human annotators.

For example, on the part-of-speech level of annotation there are currently 18,000 of those corrections, so called Bits of Wisdom (BOWs).
For applying this information to boost the NLP processing we experiment how to use the BOWs in retraining the part-of-speech tagger and found that it can be improved to correct up to 70% of identified errors within held-out data.
Moreover an improved tagger helps to raise the performance of the parser.
Preferring sentences with a high rate of verified tags in retraining has proven to be the most reliable way.

With a simulated active learning experiment using Query-by-Uncertainty (QBU) and Query-by-Committee (QBC) we proved that selectively sampling sentences for retraining yields better results with less data needed than random selection.

In an additional pilot study we found that a standard maximum-entropy part-of-speech tagger can be augmented so that it uses already known tags to enhance its tagging decisions on an entire sequence without retraining a new model first.