**1.4 Levels of NLP [read superficially]**

There are three levels at which we can “do NLP”:

**Black belt level:** reaching deep into mathematical and linguistic subtleties

**Training & tuning level:** mostly plugging in existing NLP/ML libraries

**Blackbox level:** relying on “buying” third-party NLP

**1) The black belt level**

This level comes close to computational linguistics, the academic counterpart of NLP. Two camps the mathematicians and the linguists are there. Mathematicians focus on newest methods of optimization and evaluation. Concentrate on improving the recall of your algorithms.

The linguists were raised either on highly complex generative or constraint-based grammar formalisms, or alternative frameworks such as cognitive grammar. They will gravitate towards writing syntactic and semantic rules and compiling lexica.

Depending on how you handle communication and integration between the two camps, their collaboration can either block productivity or open up exciting opportunities.

**2) The training & tuning level**

The second level involves the training and tuning of models using existing algorithms. In practice, most of the time will be spent on data preparation, training data creation and feature engineering. The core tasks — training and tuning — do not require much effort.

At this level, your people will be data scientists pushing the boundaries of open-source packages, such as nltk, scikit-learn, spacy and tensorflow, for NLP and/or machine learning. The goal is to train well-understood algorithms such as NER, categorization and sentiment analysis, customized to the specific data at your company.

The good thing here is that there are plenty of great open-source packages out there. Most of them will still leave you with enough flexibility to optimize them to your specific use case.

**3) The blackbox level**

On the third level is a “blackbox” where you buy NLP. Your developers will mostly consume paid APIs that provide the standard algorithm outputs out-of-the-box, such as Rosette, Semantria and Bitext. Ideally, your data scientists will be working alongside business analysts or subject matter experts. For example, if you are doing competitive intelligence, your business analysts will be the ones to design a model which contains your competitors, their technologies and products.

At the blackbox level, make sure you buy NLP only from black belts. With this secured, one of the obvious advantages of outsourcing NLP is that you avoid the risk of diluting your technological focus. The risk is a lack of flexibility — with time, your requirements will get more and more specific. It is also advisable to invest into manual quality assurance to make sure the API outputs deliver high quality.