Abstract: When they grow in size, knowledge bases (KBs) tend to include sets of axioms which are intuitively absurd but nonetheless logically consistent. This is particularly true of data expressed in OWL, as part of the Semantic Web framework, which favors the aggregation of set of statements from multiple sources of knowledge, with overlapping signatures.

Identifying nonsense is essential if one wants to avoid undesired inferences, but the sparse usage of negation within these datasets generally prevents the detection of such cases on a strict logical basis. And even if the KB is inconsistent, identifying the axioms responsible for the nonsense remains a non trivial task.

This thesis investigates the usage of automatically gathered linguistic evidence in order to detect and repair violations of common sense within such datasets. The main intuition consists in exploiting distributional similarity between labels of individuals of an input KB, in order to identify consequences which are unlikely to hold if the rest of the KB does. Then the repair phase consists in selecting axioms to be preferably discarded (or at least amended) in order to get rid of the nonsense.

A second strategy is also presented, which consists in strengthening the input KB with a foundational ontology, in order to obtain an inconsistency, before performing a form of knowledge base debugging/revision which incorporates this linguistic input. This last step may also be applied directly to an inconsistent input KB.

These propositions are evaluated with different sets of statements issued from the Linked Open Data cloud, as well as datasets of a higher quality, but which were automatically degraded for the evaluation. The results seem to indicate that distributional evidence may actually constitute a relevant common ground for deciding between conflicting axioms.