Despite the fact that the statistical approach is intuitively appealing, it is still prone to significant problems. Arnold (2003) attributes the deficiencies of SMT to two reasons. The first pertains to a different version of the problem of description while the second relates to the quality of the available statistical models. He states that "[T]he statistical version of the problem of description is the problem of sparse data" (ibid: 140). In reference to example 4, in order to know that *b* is more probable than *c*, huge amounts of text must be analyzed to discover that *time* appears more often that *hour* in this context. The problem is even worse when examining the aligned parallel corpora (collections of texts in two languages which are supposed to be translations of each other). Arnold (2003:139) adds "whatever the probability of seeing an expression on its own, the probability of seeing it as the translation of some other expression must generally be lower".

In her official webpage, Suliaiti, an independent researcher and an Arabic corpus consultant, lists the different Arabic corpora available. There are about 20 corpora listed, five of which are for the purpose of MT and NLP. Other corpora, which are classified on source, medium, size, purpose and material, are used for different purposes such as lexicography, pedagogy, speech recognition, etc. The list shows the scarcity of Arabic corpora in general and parallel corpora in particular as more research should be targeted at such essential field which aids many practical and applied linguistics applications.

The second hurdle to SMT Arnold presents is the statistical models (mentioned above by Manning and Shutze) utilized in the system. The standard example of a monolingual statistical model is bigram model which is used as well in speech recognition. Such a model is based on the assumption that the probability of any given word sequence can be figured out by the joint probability of each word occurring by giving the preceding word. Arnold illustrates the process as follows: