«Добродушная» точка зрения, правда, несколько эволюционировала: раньше, говоря о «прикладных задачах», подразумевали машинный перевод; теперь подразумевают задачи информационного поиска (в широком смысле).

Мы обращаем внимание читателя на то, что устойчивость указанного круга ассоциаций не является плодом случайности. Взятые в своей массе, работы по лингвистической статистике действительно дают основания для таких впечатлений и оценок. Мы, однако, не считаем, что если некто был настолько прилежен, что расписал вручную 250 тыс. слов текста, то это само по себе достойно похвалы. Тем меньше сегодня оснований превозносить успехи человека в системе «человек — машина», если все, что сделал в данном случае человек,— это составил программу, согласно которой машина «распечатает» уже не 250 тыс. слов текста, а в 100 раз больше, поскольку давно уже существуют целые библиотеки таких стандартных программ.

Тем не менее довольно много ученых полагают, что если нечто аккуратно инвентаризировано, то это и есть лингвистическая статистика. В таком случае полезно было бы понять, как сформировалось такое мнение: большинство ошибочных представлений в науке имеет достаточно серьезные основания.

По-видимому, точка зрения, согласно которой область применения СМ в лингвистике отождествляется с подсчетами частот повторяемости каких-либо единиц, порождена следующими обстоятельствами. Статистический подход применяется при исследовании совокупностей однотипных событий или объектов. Рассмотрение некоторого материала на предмет обнаружения в нем закономерностей статистического типа предполагает наличие р я д а наблюдений, в котором есть устойчивая повторяемость некоторых характеристик. Тот факт, что текст в принципе может быть рассмотрен как ряд, в котором скорее всего есть устойчивая повторяемость некоторых его элементов, был отмечен, как известно, еще в глубокой древности. К тем же временам восходит стремление к инвентаризации языковых единиц. Разнообразные частотные словари и списки в достаточно большом количестве составлялись уже начиная с середины XIX в.; при этом преследовались разнообразные практические цели типа усовершенствования системы стенографии, создания конкордансов для филологических изысканий и т. п. В первой трети XX в. уже появляются первые попытки теоретического осмысления количественных соотношений между элементами текста, хотя в целом подобные работы имели эпизодический характер вплоть до 50-х годов, когда началось бурное проникновение в лингвистику идей и методов кибернетики и теории информации. В этом русле совершенно естественным было желание выявить всевозможные количественные закономерности, наблюдаемые в текстах, словниках, в фонетическом составе языка и т. д.

Тем самым оказалось, что в лингвистике существует такая процедура, как и з м е р е н и е [6], ибо чем же еще, как не измерением, является подсчет частоты повторяемости каких-либо единиц? Возможности измерений в лингвистике резко возросли с появлением ЭВМ: теперь уже можно было не только подсчитывать частоты слов, но и автоматически выделять морфемы, словосочетания, синтаксические конструкции и считать соответствующие частоты. Накопление такого рода наблюдений сделало возможным выявление некоторых статистических закономерностей, существующих в словарях и текстах на естественном языке. Оказалось, например, что если слова любого достаточно длинного текста упорядочить по убыва-

[6] Тот тип измерений, который постоянно использовался в экспериментальной фонетике, как-то не связывался в сознании лингвистов с какими-либо методическими принципами. Считалось, что эти операции — просто следствие того, что экспериментальная фонетика — это частью акустика, частью психофизиология; значит, там применяются приборы и, следовательно, что-то регистрируется и измеряется.