

«Не нужно особого пророческого дара, чтобы предсказать: чем больше языкознание будет формализованной наукой, тем меньше оно будет наукой гуманитарной»³.

Разумеется, в наши задачи не входит обсуждение вопроса о том, как следует понимать «формализацию лингвистики». Тем не менее нам придется, пусть в несколько догматической форме, сформулировать наше понимание задач СМ в лингвистике — иначе всякое дальнейшее обсуждение будет беспредметным. Мы полагаем, что задачи СМ в лингвистике определяются следующим взглядом на задачи статистики, как таковой, высказанным известным статистиком У. Томпсоном: «Статистика — это математическая теория того, как узнать нечто о мире через опыт»⁴. Намеренная метафоричность данного высказывания позволяет нам пока не уточнять его смысл, мы сделаем это ниже.

Мы полагаем, что то место, которое СМ занимают в лингвистике сегодня, не приводит к плодотворности их для теоретической лингвистики: СМ в лингвистике применяются с избытком, но преимущественно не там, где надо, или не так, как их следовало бы применять. С другой стороны, там, где применение СМ представляется нам необходимым, они применяются недостаточно или не применяются вовсе. Обоснованию сделанных утверждений и будет в основном посвящена данная работа.

* * *

1. Что мы измеряем и с какой целью?

Рассмотрим прежде всего, какие мнения и оценки возникают у лингвиста, который в своей работе не пользуется СМ, при попытке представить себе, чем занимается «лингвистическая статистика». Пожалуй, круг ассоциаций и недоумений лингвиста можно свести к следующим трем пунктам.

1. СМ используются в лингвистике для ответа на вопрос «Как часто?». Чтобы получить такой ответ, надо обработать огромные массивы текста. После этого оказывается: полученные результаты достоверны лишь частично; кроме того, в значительной мере результаты тривиальны, ибо все это известно и без подсчетов.

2. Известно, что статистика требует привлечения большого массива текстов. Если кто-то считал вручную, то его трудолюбие оправдывается (скорее всего) целью работы. Если считали на машине, то для лингвиста, далекого от этой проблематики, возможны, как правило, две точки зрения: а) если применяли ЭВМ, значит, это хорошо, поскольку современно; б) работа с ЭВМ требует особой квалификации. Лингвист склонен думать, что его коллега, пожелавший заняться автоматической обработкой больших массивов текста, скорее всего уже давно вынужден был переквалифицироваться в вычислителя, программиста и т. п. В таком случае лингвист склонен воздержаться от комментариев — ситуация выходит за рамки его научных интересов и компетенции.

3. Во многих случаях лингвисту просто неясно, зачем нужны те или иные подсчеты. Вернее, как сказал В. И. Абаев, «я спрашиваю по простоте: а зачем мне с большой точностью знать то, что мне совсем не нужно? не лучше ли, пусть с меньшей точностью, но знать вещи действительно нужные?»⁵. Более «добродушная» точка зрения может быть выражена примерно так: «К собственно лингвистическим задачам это, конечно, отношения не имеет, но — столько сил, столько средств! — наверное, это нужно для прикладных задач».

³ Абаев В. И. Лингвистический модернизм как дегуманизация науки о языке. — Вопросы языкознания, 1965, № 3, с. 39.

⁴ Thompson W. In: The Future of Statistics, 1968; цит. по кн.: Н а л и м о в В. В. Указ. соч., с. 262.

⁵ Абаев В. И. Указ. соч., с. 32.